doi 10.4436/jass.89006

JASs *I*nvited *R*eviews
Journal of Anthropological Sciences
Vol. 89 (2011), pp. 25-58

African human mtDNA phylogeography at-a-glance

Alexandra Rosa[1,2] & António Brehm[2]

1) Medical Sciences Unit, Centre of Life Sciences, University of Madeira, Campus of Penteada, 9000-390 Funchal, Portugal
e-mail: arosa@uma.pt

2) Human Genetics Laboratory, University of Madeira, Campus of Penteada, 9000-390 Funchal, Portugal
e-mail: brehm@uma.pt

Summary - *The mitochondrial DNA (mtDNA) genetic system has long proven to be useful for studying the demographic history of our species, since their proposed Southeast/East African origin 200 kya. Despite the weak archaeological and anthropologic records, which render a difficult understanding of early intra-continental migrations, the phylogenetic L0-L1'6 split at about 140-160 kya is thought to represent also an early sub-structuring of small and isolated communities in South and East Africa. Regional variation accumulated over the following millennia, with L2 and L3 lineages arising in Central and East Africa 100-75 kya. Their sub-Saharan dispersal not later than 60 kya, largely overwhelmed the L0'1 distribution, nowadays limited to South African Khoisan and Central African Pygmies. Cyclic expansions and retractions of the equatorial forest between 40 kya and the "Last Glacial Aridity Maximum" were able to reduce the genetic diversity of modern humans. Surviving regional-specific lineages have emerged from the Sahelian refuge areas, repopulating the region and contributing to the overall West African genetic similarity. Particular L1-L3 lineages mirror the substantial population growth made possible by moister and warmer conditions of the Sahara's Wet Phase and the adoption of agriculture and iron smelting techniques. The diffusion of the farming expertise from a Central African source towards South Africa was mediated by the Bantu people 3 kya. The strong impact of their gene flow almost erased the pre-existent maternal pool. Non-L mtDNAs testify for Eurasian lineages that have enriched the African maternal pool at different timeframes: i) Near and Middle Eastern influences in Upper Palaeolithic, probably link to the spread of Afro-Asiatic languages; ii) particular lineages from West Eurasia around or after the glacial period; iii) post-glacial mtDNA signatures from the Franco-Cantabrian refugia, that have crossed the Strait of Gibraltar and iv) Eurasian lineages tracing back to the Neolithic or more recent historical episodes. Finally, the non-random sub-Saharan spread of North African lineages was likely mediated by the ancestors of Fulani, nomadic pastoral communities in the Sahel.*

Keywords - *Phylogeography, Africa, mtDNA, Polymorphisms, Population genetics.*

Introduction

Mitochondrial DNA (mtDNA) is a circular, double-stranded DNA molecule of roughly 16.6kb which encodes proteins for oxidative phosphorylation, and is present in high copy number in the energy-producing mitochondria. Together with the non-recombining portion of the Y chromosome (NRY), these constitute haploid genetic systems of uniparental inheritance, entirely and exclusively transmitted along maternal and paternal lineages, respectively. Since these molecules escape recombination, the sequential accumulation of mutations over generations are their only source of variability. Therefore, the particular features of both genetic systems have made them powerful tools for investigating the demographic history of

humankind and for addressing migratory episodes in which our maternal and paternal ancestors have participated, assuming that the present genetic variability reflects past demographic events. Conventionally, haplotypes are defined by the entire profile of mutations along the mtDNA molecule comparatively to a consensus sequence (Cambridge Reference Sequence, CRS; Anderson *et al.*, 1981, revised by Andrews *et al.*, 1999). The variants sharing a common ancestor (extant or reconstructed) cluster hierarchically in clades with common mutations – called haplogroups. Phylogenetic relationships of known mtDNA variants are illustrated by means of phylogenetic trees or networks.

The research of mtDNA as a phylogenetic molecular marker was pioneered by Wesley Brown and Douglas Wallace in the late 1970s (Brown, 1980). Their work hit the spotlight by evidencing a common evolutionary history for our species, given that the global phylogeny of present-day populations inhabiting the four corners of the globe was found to coalesce in a single central node. Early PCR-RFLP studies were based on the mtDNA coding region (spanning nucleotide positions, nps 577-16023), where homoplasy is rarer and therefore more stable (e.g. Denaro *et al.*, 1981; Johnson *et al.*, 1983; Cann *et al.*, 1987; Scozzari *et al.*, 1988; Soodyall & Jenkins, 1992, 1993; Torroni *et al.*, 1992). Subsequent analyses adopted a more informative synthesis of coding region and first hypervariable segment (HVS-I, nps 16024-16365) mutation patterns (Torroni *et al.*, 1996; Macaulay *et al.*, 1999; Bandelt *et al.*, 2000; Chen *et al.*, 2000; Kivisild *et al.*, 2002; Jobling *et al.*, 2004). While the basal topology of an mtDNA phylogeny is overwhelmingly based on coding region SNPs, HVS-I information resolves independent lineages and branches at the tips of the topology, following the principle of parsimony (*i.e.* the least evolutionary changes). More importantly, many of the phylogenetic clades were found to be geographically structured, a door-opening assertion for the approach known as phylogeography (Avise, 2000).

Unequivocal phylogenetic construction using appropriate methods (Bandelt *et al.*, 1995, 1999, 2000) makes it possible to estimate coalescence ages of clades in the phylogeny and therefore, to reconstruct and interpret relevant demographic events. In the words of Bandelt, Macaulay and Richards "only when reconstructed and dated ancestral types appear to have given rise to essentially autochthonous branches of the phylogeny, with approximately equal coalescence time, then one could speak of founder types at the colonizing event" (Bandelt *et al.*, 2006). The inclusion of a timeframe relies on the accurate estimation of the mutation rate, and thus of a well calibrated molecular clock, associating an absolute timescale to the sequence diversity. Ever since the estimation of 20180 years *per* mutation on the HVS-I nps 16090-16365 stretch (1.26 x 10^{-8} substitutions/site/year; Forster *et al.*, 1996, 2000), this has been a topic of open debate in literature (e.g. Macaulay *et al.*, 1997, 2005; Meyer *et al.*, 1999; Ingman *et al.*, 2000; Sigurgardottir *et al.*, 2000; Heyer *et al.*, 2001; Maca-Meyer *et al.*, 2001; Howell *et al.*, 2003, 2004; Bandelt *et al.*, 2006; Kivisild *et al.*, 2006). The 1.79 x 10^{-7} and 3.5 x 10^{-8} substitutions/site/year on coding region nps 577-16023 and synonymous changes at protein coding sites, respectively suggested by Mishmar *et al.* (2003) and Kivisild *et al.* (2006), are nowadays among the most widely used molecular clocks to translate age estimations in mutations into age estimates in years, *i.e.* the coalescence time to the Most Recent Common Ancestor (MRCA) of a set of mtDNA lineages.

A critical assessment of methodologies and modes of calibration of the mtDNA molecular clock was recently performed in Endicott *et al.* (2009). Among the main assumptions found to be of contestable validity are *i)* the reliance on human-chimpanzee genetic split at 6 Mya and *ii)* the linear, selection-independent accumulation of substitutions in time, since the split of these species to the present. A higher proportion of synonymous mutations in younger branches, a likely reflex of purifying selection (or the relaxation of selective constraints), has in fact been already stated by several authors (e.g. Kivisild *et al.*, 2006; Soares *et al.*, 2009). A *post-hoc* correction formula for the modest effect of purifying

selection on the mtDNA coding region and the effect of time depth on mutation rate has been recently proposed by Soares *et al.* (2009). Nevertheless, the method does not allow for rate variation among contemporaneous lineages. In the present state-of-the-art, our current interpretation of the evolutionary timescale and demographic history of our species based on genetic data is threatened. Besides the movement away from human-chimp calibration (and towards the within-human mtDNA tree calibration), Endicott *et al.* (2009) recommend the Bayesian phylogenetic approaches which, within a single framework, implement models that allow for rate heterogeneity among sites and among lineages, further correcting for multiple hits.

For the purpose of comparing worldwide lineages, a universal nomenclature system was required. The study of Native Americans by Torroni *et al.* (1993) initiated the currently accepted nomenclature, by describing four basal Native-American clusters in alphabetical order – haplogroups A, B, C and D. The classification has nevertheless become a continuous process where new data emerge within short periods of time, allowing frequent adjustments to a better phylogenetic resolution. To the capital letters representing the haplogroups, subclusters were attributed additional symbols (alternating letters and numbers, e.g. L0a1). An asterisk symbol (*) is used to refer to paragroups, different yet unidentified clades, that can even be MRCAs. Several attempts have been made to homogenize the nomenclature (Richards *et al.*, 1998; Kivisild *et al.*, 2006).

High-throughput automatic sequencing technologies have led to a boom in publications on mtDNA full-molecule sequences in particular (e.g. Finnilä *et al.*, 2001; Torroni *et al.*, 2001b; Kivisild *et al.*, 2002, 2006; Ingman & Gyllensten, 2003; Mishmar *et al.*, 2003; Achilli *et al.*, 2004, 2005; Macaulay *et al.*, 2005; Gonder *et al.*, 2006; Olivieri *et al.*, 2006; Underhill & Kivisild, 2007). Trees of complete mtDNA genomes allow a topological refinement, with dissection of earlier unresolvable haplogroups (e.g. the typically European haplogroup H; Achilli *et al.*, 2004; Loogväli *et al.*, 2004; Pereira *et al.*, 2006; Roostalu *et al.*, 2007) and more precise estimates of temporal layers. In fact, and given the many fields of application of mtDNA, namely evolutionary anthropology, population history, genetic genealogy, and forensic and medical genetics, much of the present work makes use of complete mtDNA sequences. Therefore, there was an urgent demand for a detailed knowledge of the phylogenetic relationships of globally described mtDNA variants. An immense amount of information (nearly 4,200 complete mtDNA genomes) was recently compiled in a comprehensive and frequently updated phylogeny by van Oven & Kayser (2009), publicly available online at www.phylotree.org. Previous attempts to formulate a global mtDNA phylogeny were less successful given the limited number of samples and quick outdating of the information, the absence of control region information and/or absence of detailed nomenclature (e.g. MITOMAP, mtDB; Ingman & Gyllensten, 2006; Ruiz-Pesini *et al.*, 2007). Other tools such as Mitomaster (Brandon *et al.*, 2009) and GeneSyn (Pereira *et al.*, 2009) have been developed to exhaustively analyze the diversity in genetic data sets, among other variables considering the synonymous and non-synonymous character of the polymorphisms, their conservation degree among species, chemical and physical properties of the amino acids, and the population frequencies as part of the evaluation of their pathogenic effect.

The African origin and first dispersals of Anatomically Modern Humans (AMH)

Although initially criticised, Allan Wilson's group was among the first to consistently suggest an African origin for all humankind maternal lineages (Cann *et al.*, 1987; Vigilant *et al.*, 1991). The phylogenetic relationships of worldwide samples showed a deep basal split between a clearly exclusive African branch and one other encompassing the remaining African and non-African types. Furthermore, it has been observed that African

mtDNA lineages had the highest known diversity, implicating that the ancestral variant (also known as "mitochondrial Eve") has been present in Africa much earlier than elsewhere. If one assumes that mtDNA types in the present human gene pool can ultimately be traced to a single common ancestor, the maternal lineages of all living humans coalesce at a time depth of 160-200 kya, in a Southeast or East African cradle (Watson *et al.*, 1997; Kivisild *et al.*, 1999; Ingman *et al.*, 2000; Maca-Meyer *et al.*, 2001; Mishmar *et al.*, 2003; Forster, 2004; Macaulay *et al.*, 2005; Gonder *et al.*, 2006; Torroni *et al.*, 2006; Behar *et al.*, 2008). The timeframe is coincident with paleontological data, currently found for the emergence of early AMH (Day *et al.*, 1982; Rightmire, 1989; Grun *et al.*, 1990; Foley, 1998; Rightmire & Deacon, 2001; White *et al.*, 2003; McDougall *et al.*, 2005; Rightmire *et al.*, 2006).

Furthermore, two antagonistic models were in the highlights to explain the non-African variation: the "multiregionalism" versus the "recent Out-of-Africa". mtDNA evidence indeed supported the "Out-of-Africa" scenario (e.g. Horai, 1995; Jorde *et al.*, 1995; Watson *et al.*, 1997; Ingman *et al.*, 2000; Forster, 2004), with a AMH origin in Africa, where evolution proceeded for 80-120 kya, and from where it spread over other regions completely replacing regionally different earlier humans (Lewin, 1987; Stringer & Andrews, 1988; Foley, 1998; Stringer, 2000, 2003).

Among the first emerging subsets of our maternal variation, at about 150-180 kya and which have reached the present day, are haplogroup L0 and the branch encompassing L1'L6 and all the non-African variation. Additional variation accumulated over the following millennia in parallel to migratory events and periods of isolation (Forster, 2004), mostly due to climatic and cultural constraints as evidenced by palaeoclimatology and archaeology (Lahr & Foley, 1994, 1998; Henshilwood *et al.*, 2002; Mellars & Crow, 2002). A fragmented environment and elements of modern human behaviour in several parts of Africa were dated not later than 70 kya, which is temporally coincident with the emergence of L2 lineages in Central Africa (about 90-105 kya; (Gonder *et al.*, 2006; Behar *et al.*, 2008; Soares *et al.*, 2009). Haplogroup L3 arose most probably within the eastern African mtDNA pool, at about 65-80 kya and subsequently diverged *in situ* into a multitude of subclades (Salas *et al.*, 2002; Forster, 2004; Kivisild *et al.*, 2004, 2006; Macaulay *et al.*, 2005; Torroni *et al.*, 2006; Behar *et al.*, 2008; Atkinson *et al.*, 2009). Out of this plethora of African-specific L3 lineages, just two – M and N (~60-65 kya; Forster *et al.*, 2001; Kong *et al.*, 2003; Mishmar *et al.*, 2003; Macaulay *et al.*, 2005) – have successfully made their way out of Africa, giving rise to the mtDNA pool of all non-Africans. As mentioned by Gonder and co-workers, it is worth pointing out that the coalescence estimate for L3, M and N derivatives is of nearly half the ultimate coalescence of all AMHs, therefore reinforcing the exclusive within-Africa evolution of AMH mtDNA lineages for roughly 100 ky, before the exodus to other regions (Gonder *et al.*, 2006). The question of whether the two M and N macrohaplogroups that colonized Eurasia were already present in Africa before the exit has yet to be settled (Kivisild *et al.*, 2003, 2004; Gonder *et al.*, 2006).

Extant mtDNA lineages in the African continent

The following paper intends to summarize the present-day knowledge about African mtDNA haplogroups (and their subclusters), their proposed origins, coalescence ages and relevant demographic and migratory events, which can be traced on a genetic basis. Note that coalescence ages throughout the text are based on HVS-I and/or coding region/full mtDNA information (with appropriate references), and therefore not necessarily coincident. A summary of selected works on HVS-I (Salas *et al.,* 2002) and coding region/full molecule estimates (Behar *et al.*, 2008; Soares *et al.*, 2009) for main extant African haplogroups is shown in Table 1.

It is relevant to note that most mtDNA haplogroups in the African continent show variable distributions and sub-structuring when

geography and ethnolinguistic affiliations are considered (Watson *et al.*, 1997; Chen *et al.*, 2000; Pereira *et al.*, 2001; Salas *et al.*, 2002; Destro-Bisol *et al.*, 2004; González *et al.*, 2006).

Haplogroup L0

Macrohaplogroup L divides into haplogroups L0-L6 (Salas *et al.*, 2002, 2004; Mishmar *et al.*, 2003; Kivisild *et al.*, 2006) and is mainly limited to sub-Saharan Africa. Haplogroup L0 turned out to be one of the earliest offshoots of the mtDNA variation at about 140-160 kya, a sister clade of that holding all other AMH extant mtDNA haplogroups (Mishmar *et al.*, 2003; Gonder *et al.*, 2006; Torroni *et al.*, 2006; Behar *et al.*, 2008, Soares *et al.*, 2009) (Fig.1). Haplogroup L0 further includes sub-haplogroups L0a, L0d, L0f and L0k (Salas *et al.*, 2004; Gonder *et al.*, 2006; Kivisild *et al.*, 2006; Behar *et al.*, 2008). L0d, the first individual sub-clade to derive from the L0 node, displays an estimated coalescence age of 100 kya, although its divergence from L0abfk probably started 144 kya (Gonder *et al.*, 2006; Kivisild *et al.*, 2006; Behar *et al.*, 2008) (Fig.1). The distribution of this clade appears to be restricted to Khoisan people in South Africa, and to Tanzanian and Angolan populations (Vigilant *et al.*, 1991; Chen *et al.*, 2000; Pereira *et al.*, 2001; Salas *et al.*, 2002; Kivisild *et al.*, 2004; Gonder *et al.*, 2006; Tishkoff *et al.*, 2007; Coelho *et al.*, 2009, Figure 2). The most recently accepted tree topology makes it possible to identify L0k as a sister clade to that including L0abf (Gonder *et al.*, 2006; Kivisild *et al.*, 2006; Behar *et al.*, 2008). Similarly, sub-haplogroup L0k is found almost exclusively among South African Khoisan (Chen *et al.*, 2000; Salas *et al.*, 2002) existing also at low frequencies among click-speaking Tanzanian groups (Tishkoff *et al.*, 2007). Their L0d and L0k shared lineages, which represent more than half of their maternal pool (Fig.2), suggest an ancestral link predating the appearance of present-day click-speakers, likely remnants of an East African proto-Khoisan population (Bandelt & Forster, 1997; Chen *et al.*, 2000; Pereira *et al.*, 2001; Salas *et al.*, 2002; Gonder *et al.*, 2006; Tishkoff *et al.*, 2007). L0k lineages are dated by Behar *et al.* (2008) at about 40 kya (Fig.1). The rare L0f lineages are present and more diverse only in East Africa, where they likely arose 85-90 kya (Fig.1), with their highest incidence in Tanzanians (Knight *et al.*, 2003; Brandstätter *et al.*, 2004; Tishkoff *et al.*, 2007; Behar *et al.*, 2008; Castri *et al.*, 2009). L0a lineages probably originated in eastern Africa in Paleolithic times at about 40-55 kya (Watson *et al.*, 1997; Salas *et al.*, 2002; Behar *et al.*, 2008; Soares *et al.*, 2009) and are today widely spread through eastern, central and southern Africa, in some cases encompassing more than a quarter of maternal lineages there (Watson *et al.*, 1997; Chen *et al.*, 2000; Salas *et al.*, 2002, 2004; Kivisild *et al.*, 2004; Coia *et al.*, 2005; Gonder *et al.*, 2006; Tishkoff *et al.*, 2007; Quintana-Murci *et al.*, 2008; Castri *et al.*, 2009; Coelho *et al.*, 2009). The L0a1 sub-clade has an eastern and southeastern African distribution including Nubia, Sudan and Ethiopia, with the root type coalescing at nearly 30 kya (Salas *et al.*, 2002; Behar *et al.*, 2008; Soares *et al.*, 2009). At rather low frequencies, L0a1 is also found in West Africa (Rando *et al.*, 1998; Brehm *et al.*, 2002; Arredi *et al.*, 2004; Jackson *et al.*, 2005; Cerný *et al.*, 2006; Ely *et al.*, 2006). The phylogenetic picture for L0a is characterized by several short branches, suggesting recent population growth. Together with their distribution pattern, specific L0a2 lineages are thought to trace the dispersal of Bantu-speakers towards South Africa 3 kya (Soodyall *et al.*, 1996; see further details in the following sections).

Haplogroup L1

MtDNA L1 lineages are found to coalesce at about 140-150 kya (Torroni *et al.*, 2006; Behar *et al.*, 2008) (Fig.1). One of its daughter clades, haplogroup L1b, is concentrated in western-central Africa, particularly along the coastal areas (Watson *et al.*, 1997; Rando *et al.*, 1998; Brehm *et al.*, 2002; Rosa *et al.*, 2004; Coia *et al.*, 2005; González *et al.*, 2006), peaking in the Senegal Mandenka and Wolof (Rando *et al.*, 1998; Jackson *et al.*, 2005) and Fulani people in Burkina-Faso, Chad and South Cameroon (Cerný *et al.*, 2006) (Fig.2). The extant variation of L1b probably mirrors a

Tab. 1 - Coalescence estimates for L0-6 mtDNA haplogroups (expressed in kya +/- standard deviation).

	SALAS *ET AL.* 2002[A]	BEHAR *ET AL.* 2008[B]	SOARES *ET AL.* 2009[C]
L0	ND	153,971+/-11,905	149,700 (112,200-188,000)
L0a	40,350+/-16,250	53,176+/-7,143	44,800 (28,900-61,500)
L0a1	33,350+/-16,600	26,191+/-5,556	26,800 (15,100-39,100)
L0d	49,600+/-13,450	101,589+/-10,318	ND
L0k	ND	39,683+/-8,730	11,200 (3,300-19,500)
L0f	ND	88,097+/-9,524	ND
L1b	30,550+/-16,250	29,366+/-7,937	9,700 (5,200-14,300)
L1c	59,650+/-11,800	102,383+/-7,937	85,400 (63,500-107,900)
L5	ND	138,098+/-11,905	120,200 (114,200-126,200)
L2	70,100+/-15,300	104,764+/-8,730	89,300 (68,100-111,100)
L2a	55,150+/-19,350	46,033+/-7,937	48,300 (31,100-66,300)
L2b	31,600+/-11,200	26,985+/-3,968	14,300 (7,000-21,900)
L2c	27,500+/-7,250	23,810+/-3,175	25,200 (13,400-37,700)
L2d	121,900+/-34,200	23,810+/-7,143	ND
L2e	ND	46,826+/-7,143	ND
L6	ND	22,223+/-5,556	ND
L3	61,300+/-11,650	76,192+/-4,762	71,600 (57,100-86,600)
L3b	21,600+/-6,850	28,572+/-4,762	16,400 (9,800-23,200)

clear-cut example of a bottleneck that has shaped its evolution, leaving no other progeny except a clade that began expanding at about 30 kya (Salas *et al.*, 2002; Kivisild *et al.*, 2004; Rosa *et al.*, 2004; Behar *et al.*, 2008).

Its sister clade L1c occurs frequently in Central and West Africans (Rando *et al.*, 1998; Brehm *et al.*, 2002; Cerný *et al.*, 2004; Rosa *et al.*, 2004; Coia *et al.*, 2005; Jackson *et al.*, 2005; González *et al.*, 2006; Batini *et al.*, 2007), representing over 70% of the maternal legacy of many Pygmy groups (Watson *et al.*, 1997; Destro-Bisol *et al.*, 2004; Quintana-Murci *et al.*, 2008) (Fig.2). Curiously, more recent reports state frequencies ranging 18-25% in Angola Bantu ethnic groups (Beleza *et al.*, 2005; Coelho *et al.*, 2009). A substantial revision for the L1c phylogeny has been proposed by Quintana-Murci *et al.* (2008). It shed additional light over the cultural transition from hunting-gathering to agriculture and helped corroborate past relationships between Central African Bantu-speaking farmers and their hunter-gathering neighbors, the Pygmies, as already suggested (Salas *et al.*, 2002; Batini *et al.*, 2007). Both groups likely shared an ancestral Central-African proto-population rich in L1c mtDNAs, which started to diverge in isolation not later than 70 kya, and evolved into the diverse forms observed today among the modern agricultural populations (L1c1a, L1c1b, L1c1c, L1c2-6, etc.) while L1c1a is the only surviving clade in western Pygmies (Quintana-Murci *et al.*, 2008).

Tab. 1 - Continued

	SALAS *ET AL.* 2002[A]	BEHAR *ET AL.* 2008[B]	SOARES *ET AL.* 2009[C]
L3d	30,250+/-8,450	38,096+/-5,556	31,000 (19,200-43,300)
L3e	49,250+/-11,750	44,445+/-5,556	39,000 (27,200-51,200)
L3e1	32,150+/-11,450	15,873+/-2,381	ND
L3e2	37,400+/-18,350	26,985+/-7,143	ND
L3e3	14,150+/-4,500	38,096+/-5,556	ND
L3e4	24,200+/-10,400	ND	ND
L3f	36,300+/-12,800	60,319+/-6,349	53,200 (38,600-68,300)
L3f1	28,650+/-8,650	51,588+/-6,349	ND
L3f2	ND	60,319+/-6,349	ND
L3h	ND	70,636+/-5,556	66,700 (52,000-81,800)
L4	ND	95,240+/-7,143	ND
L4a	ND	54,763+/-7,143	ND
L4b	45,100+/-12,500	90,478+/-7,937	ND

ND - not determined.
Coalescence estimates are based on:
[a] - HVS-I sequence information, Forster et al. (1996) molecular clock;
[b] - full mtDNA sequence information, Mishmar et al. (2003) molecular clock;
[c] - mtDNA coding region, Soares et al. (2009) molecular clock.

Interestingly, L1c lineages also made it possible to trace a later gene flow between ancestors of both groups at about 40 kya (Quintana-Murci *et al.*, 2008). Both L1b and L1c were proposed as Central Africa autochthonous lineages, the latter with about 85-100 ky (Gonder *et al.*, 2006; Behar *et al.*, 2008; Soares *et al.*, 2009) (Fig.1), their presence on the West Atlantic coast suggesting a westwards expansion. Their diffusion to Northwest Africa was probably more recent, in the Neolithic or during the times of the slave trade (Rando *et al.*, 1998; Rosa *et al.*, 2004; Cherni *et al.*, 2009).

Haplogroup L5

Haplogroup L5, previously known as L1e, occupies an intermediate position between L1 and L2'3'4'6 and is probably 120-140 ky old (Kivisild *et al.*, 2004, 2006; Gonder *et al.*, 2006; Torroni *et al.*, 2006; Behar *et al.*, 2008; Soares *et al.*, 2009) (Fig.1). It has been observed at low frequencies in eastern Africa, namely Egypt, Sudan, Ethiopia, Kenya, Rwanda and Tanzania, with minor gene flow introducing these lineages in the Mbuti Pygmies and North Cameroon Fali (Krings *et al.*, 1999; Brandstätter *et al.*, 2004; Kivisild *et al.*, 2004; Stevanovitch *et al.*, 2004; Coia *et al.*, 2005; Tishkoff *et al.*, 2007; Castri *et al.*, 2009). The Central African Pygmies particular genetic pool including both L1c and L5 may assign them a "relict" status, similar to that proposed for the Khoisan (Chen *et al.*, 2000; Gonder *et al.*, 2006; Quintana-Murci *et al.*, 2008).

Haplogroup L2

Together with L3, haplogroup L2 comprises ~70% of the sub-Saharan maternal variation.

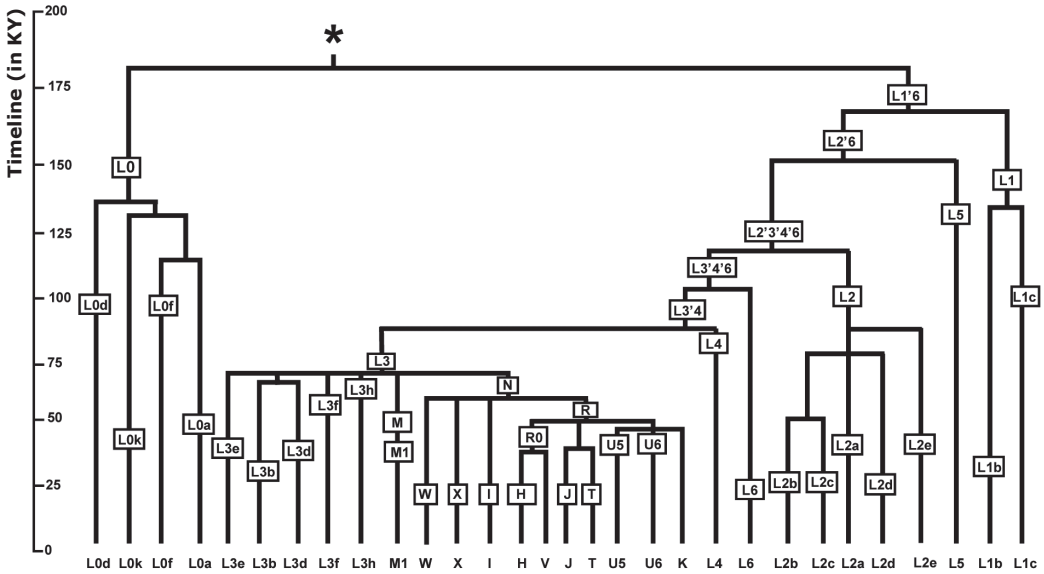

Fig. 1 - Phylogenetic tree of mtDNA haplogroups with African distribution and their estimated coalescence ages. Based on Behar et al. (2008) for L0-6 haplogroups and references in the text for M1, N and R clades.

Chen *et al.* (2000) and Torroni *et al.* (2001) were among the first to dissect haplogroup L2 into sub-clades L2a, L2b, L2c, L2d and L2e. Haplogroup L2a is the most frequent and widespread mtDNA cluster in Africa, reaching over 40% in Tuareg from Niger/Nigeria and Mali (Watson *et al.*, 1997, Salas *et al.*, 2002), Fali from North Cameroon (Coia *et al.*, 2005), Western Pygmies from Gabon (Quintana-Murci *et al.*, 2008) and Mozambique Bantu (Pereira *et al.*, 2001) (Fig.2), which makes its geographic origin difficult to identify. Furthermore, the diversity patterns of L2a sub-clades and their coalescence time at about 45-55 kya are comparable in West and East Africa (Salas *et al.*, 2002; Kivisild *et al.*, 2004; Rosa *et al.*, 2004, Behar *et al.*, 2008; Soares *et al.*, 2009) (Fig.1). A putative explanation considers its origin in more central areas, followed by west- and eastwards dispersals along the Sahel corridor after the LGAM ("Last Glacial Aridity Maximum", the African phenomenon homologous to the Eurasiatic LGM) as shared founder types coalesce at about 14 kya (Salas *et al.*, 2002; Rosa *et al.*, 2004). Recent star-like demographic bursts in L2a1a and L2a2 and their expansion to southeast people are most likely associated with the expansion(s) of the Bantu-speaking populations during the sub-Saharan agricultural spread and later on (Pereira *et al.*, 2001; Salas *et al.*, 2002; Atkinson *et al.*, 2009).

L2b-L2d haplogroups are dominant and largely confined to West and West-Central Africa, probably their region of origin (Vigilant *et al.*, 1991; Rando *et al.*, 1998; Rosa *et al.*, 2004; Coia *et al.*, 2005; Jackson *et al.*, 2005; Cerný *et al.*, 2006; González *et al.*, 2006) (Fig.2). The coalescence estimate for L2b and L2c is about 25-30 kya (Salas *et al.*, 2002; Rosa *et al.*, 2004, Behar *et al.*, 2008) (Fig.1). Nevertheless, HVS-I estimates of specific L2b and L2c West African lineages at about 18 kya (Rosa *et al.*, 2004) indicates an expansion paralleling that of L1b in West Africa (Rando *et al.*, 1998). L2d is a West/Central African infrequent clade that diverged from the L2 common root at approximately 100-120 kya (Salas *et al.*, 2002; Rosa *et al.*, 2004; Behar *et al.*, 2008) but whose

extant variation is not older than 25-30 ky (Behar *et al.*, 2008) (Fig.1). Given the timeframe considered for its origin and the absence of founder lineages in West Africa (Rosa *et al.*, 2004), the data are more consistent with its Central African origin.

Haplogroup L6

The variation classified as haplogroup L6, whose labeling was proposed by Kivisild *et al.* (2004), is nowadays largely confined to Yemeni people and a few samples in Ethiopian Amhara and Gurages. It is noteworthy that L6 presents a very narrow phylogeography for the *ca.* 110 ky divergence from its L3'4 sister clades (Torroni *et al.*, 2006). However, its own coalescence time is only about 22 kya (Behar *et al.*, 2008) (Fig.1), presumably because past variation was wiped out or actually never expanded. Given its presence in Ethiopians, where its sister clades are also diverse and frequent (Kivisild *et al.*, 2004), L6 has a most likely origin in East Africa, where it might have been preserved in isolation for tens of thousands of years. In any case the homeland of L6 may still be missing.

Haplogroup L4

Haplogroup L4 is a sister clade of L3, typical of East and Northeast Africa, although present at low frequencies (Watson *et al.*, 1997; Krings *et al.*, 1999; Kivisild *et al.*, 2004; Tishkoff *et al.*, 2007; Castri *et al.*, 2009). Here, we drop the label L7 in the review of Torroni *et al.* (2006) and use the original L4a label according to the reasons presented in Behar *et al.* (2008). The L4a motif has been found in Sudan and Ethiopia, though initially misclassified as L3e4 in Salas *et al.* (2002). Similarly we also refer to L4b2, previously known as L3g (Salas *et al.*, 2002) or L4g (Kivisild *et al.*, 2004) but more recently renamed by Behar *et al.* (2008). This is frequent in Tanzania and Amhara and Gurages from Ethiopia (Salas *et al.*, 2002; Kivisild *et al.*, 2004; Gonder *et al.*, 2006). L4a and L4b variation are estimated to coalesce at about 55 and 90 kya, respectively, while the all L4 clade probably arose 95 ky (Behar *et al.*, 2008) (Fig.1) possibly in East Africa (Kivisild *et al.*, 2004).

Haplogroup L3

An East African origin at about 60-75 kya is generally attributed to superhaplogroup L3 (Salas *et al.*, 2002; Macaulay *et al.*, 2005; Kivisild *et al.*, 2006; Torroni *et al.*, 2006; Behar *et al.*, 2008; Soares *et al.*, 2009) (Fig.1). L3 is widespread in Africa, its frequency and diversity providing evidence of a sub-Saharan expansion of its sub-clades towards West Africa (Watson *et al.*, 1997; Salas *et al.*, 2002). This superhaplogroup is subdivided into various clades and harbours also the two main M and L superhaplogroups found outside of Africa.

Both L3b and L3d are prevalent in the West quadrant of sub-Saharan Africa (in average 10%; Rando *et al.*, 1998; Brehm *et al.*, 2002; Cerný *et al.*, 2004; Rosa *et al.*, 2004; Coia *et al.*, 2005; Jackson *et al.*, 2005; Cerný *et al.*, 2006; González *et al.*, 2006) (Fig.2). L3b also shows considerable frequencies in the Hutu people in Rwanda (Castri *et al.*, 2009) and South African Kung (Chen *et al.*, 2000). L3d constitutes an important percentage of the South African maternal pool, being more expressive in Angola and Tanzania (Tishkoff *et al.*, 2007; Coelho *et al.*, 2009). Their split from a common node is estimated at about 70 kya, while their coalescence is about 20-30 kya for L3b and 30-40 kya for L3d (Salas *et al.*, 2002; Torroni *et al.*, 2006; Behar *et al.*, 2008; Soares *et al.*, 2009) (Fig.1). As already stated for other haplogroups, a subset of L3b is common among Bantu speakers of south-western Africa and thus is a likely marker of the Bantu expansion (Watson *et al.*, 1997) (Fig.2). The L3e cluster has been subdivided into L3e1, L3e2, L3e3 and L3e4, since the time of HVS-I information *per se* (Bandelt *et al.*, 2001). The oldest branches of L3e are thought to have arisen in Central Africa/nowadays Sudan *ca.* 40-50 kya (Salas *et al.*, 2002; Rosa *et al.*, 2004; Torroni *et al.*, 2006; Behar *et al.*, 2008; Soares *et al.*, 2009) (Figs.1 and 2). From here, they spread throughout sub-Saharan Africa, comprising by now about one third of L3 types (Watson *et al.*, 1997; Rando *et al.*, 1998; Pereira *et al.*, 2001; Brehm *et al.*, 2002; Salas *et al.*, 2002; Cerný *et al.*, 2004, 2006; Rosa *et al.*, 2004; Trovoada *et al.*, 2004; Coia *et al.*, 2005; Jackson *et al.*, 2005;

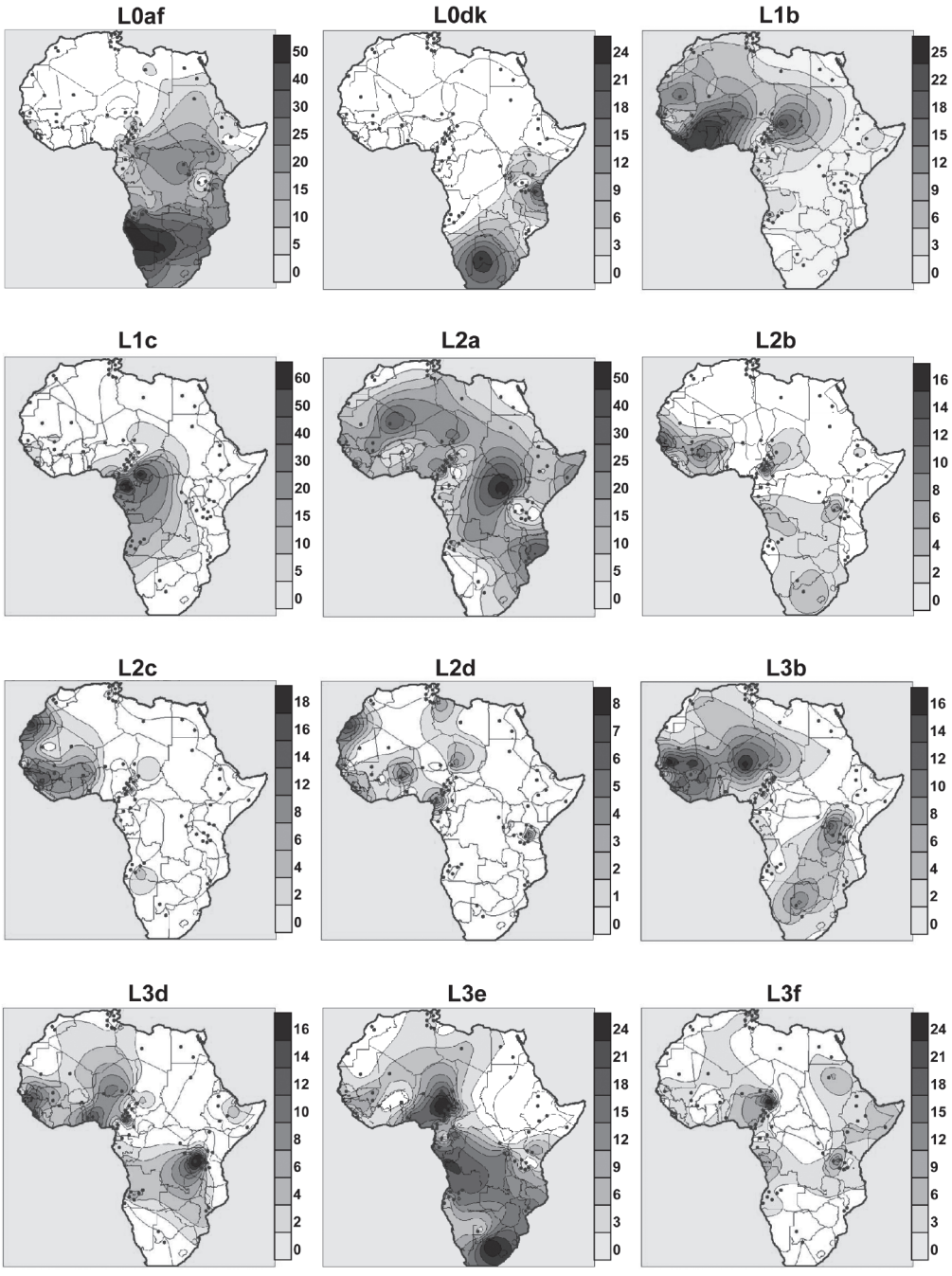

Fig. 2 - Frequency distribution maps for L0af, L0dk, L1b, L1c, L2a, L2b, L2c, L2d, L3b, L3d, L3e, L3f haplogroups present in Africa. Maps were generated using the Surfer 8.02 software and redrawn by hand. Populations used to generate these frequency clines are shown in Appendix.

González *et al.*, 2006; Quintana-Murci *et al.*, 2008; Coelho *et al.*, 2009) (Fig.2). Nevertheless, their presence in North Africa at an average of 5% cannot be disregarded (Rando *et al.*, 1998; Cherni *et al.*, 2009; Ottoni *et al.*, 2009). L3e1 probably arose at about 16 kya (Behar *et al.*, 2008) in the central areas of the continent and became quite frequent among southeast Bantu-speakers, especially in Mozambique (Bandelt *et al.*, 2001; Pereira *et al.*, 2001; Salas *et al.*, 2002), which may be linked to an eastern Bantu route. Within L3e2, the L3e2b lineages constitute the most frequent and widespread type of L3e, primarily found in West and Central Africa (Watson *et al.*, 1997; Salas *et al.*, 2002; Rosa *et al.*, 2004; Jackson *et al.*, 2005). Together with L3e2a, and despite their most recent common ancestor arose at about 25-35 kya (Salas *et al.*, 2002; Rosa *et al.*, 2004; Behar *et al.*, 2008), their range expansion at about 9 kya (Salas *et al.*, 2002) highlights these lineages as successful hitchhikers of population movements in the Sahara during the Great Wet Phase of the early Holocene and subsequent Wet Phase (Muzzolini *et al.,* 1993; Bandelt *et al.,* 2001). L3e3 is found primarily in West African people, where the variation coalesces at 14 kya (Salas *et al.*, 2002). The root type and a few derivatives are also observed in southeast Africans, raising a possible connection with the eastern stream of Bantu people. L3e4 has an age of nearly 24 ky (Salas *et al.*, 2002) and is also essentially restricted to Atlantic West Africa, signalling dispersals and local expansion events with the rise of food production and iron smelting (Bandelt *et al.*, 2001; Rosa *et al.*, 2004). A subset of L3e sequences has been named as L3e5 by Cerný *et al.* (2007). Although similar sequences were earlier detected in Tunisian Berbers (Fadhlaoui-Zid *et al.*, 2004), these were left unnamed and suggested to be of North African origin. The network in Cerný *et al.* (2007) reflects a clear star-like phylogeny of L3e5 types found mostly in western Central Africa. Although an important diffusion has occurred into North Africa, the root type is relatively prevalent in the Chad Basin populations where their expansion is estimated at about 12 kya.

The diffusion of haplogroup L3f ranges from Ethiopia in the east, to Angola and Mozambique in the south, the Chad Basin in Central Africa, Guinea-Bissau in the west and Tunisia in the north (Watson *et al.*, 1997; Rando *et al.*, 1998; Krings *et al.*, 1999; Salas *et al.*, 2002; Cerný *et al.*, 2004; Kivisild *et al.*, 2004; Rosa *et al.*, 2004; Beleza *et al.*, 2005; Coia *et al.*, 2005; Jackson *et al.*, 2005; Tishkoff *et al.*, 2007; Quintana-Murci *et al.*, 2008; Cherni *et al.*, 2009; Coelho *et al.*, 2009) (Fig.2). Coalescence estimates and a few matches to L3f1 founder lineages in Central and West Africa (Salas *et al.*, 2002; Rosa *et al.*, 2004) point to a local early dispersal of the lineages (at about 30 kya if based on HVS-I estimates in Salas *et al.*, 2002 and Rosa *et al.*, 2004 or 50 kya if based on full mtDNA data of Behar *et al.*, 2008) while in East Africa, L3f1 only started to expand ~10 kya (Kivisild *et al.*, 2004). On the other hand, L3f2 is a quite infrequent clade found almost exclusively among Chading-speaking populations from the Chad Basin and virtually absent from Niger-Congo and Nilo-Saharan peoples (Cerný *et al.*, 2007, 2009). Its MRCA is estimated at about 60 ky old (full mtDNA estimate of Behar *et al.* 2008; or 29 ky if based on HVS-I only; Cerný *et al.*, 2007, 2009). Therefore, it is contemporary with its sister clade L3f1, and probably arose around the Chad Basin area. Nevertheless, the haplogroup is present in northern Cushitic groups from Somalia and Ethiopia (Watson *et al.*, 1997; Kivisild *et al.*, 2004), suggesting the migration of proto-Chadic pastoralists from East/Northeast Africa to the Chad Basin, probably during the Holocene 8kya (Cerný *et al.*, 2009).

The subset of variation characterizing L3h lineages (more specifically L3h1b) was first labelled by our team in the context of a HVS-I-coding region RFLPs survey of Guinean populations (Rosa *et al.*, 2004). Close HVS-I variants are described in Cape Verdeans (Brehm *et al.*, 2002) and Ethiopian Amharans (Kivisild *et al.*, 2004) at low frequencies (~1%), and reach their highest known frequency in the Zriba in Tunisia (Cherni *et al.*, 2009), Ejamat in Guinea-Bissau (Rosa *et al.*, 2004) and Datoga people in Tanzania (Tishkoff *et al.*, 2007) (Fig.3). However, caution

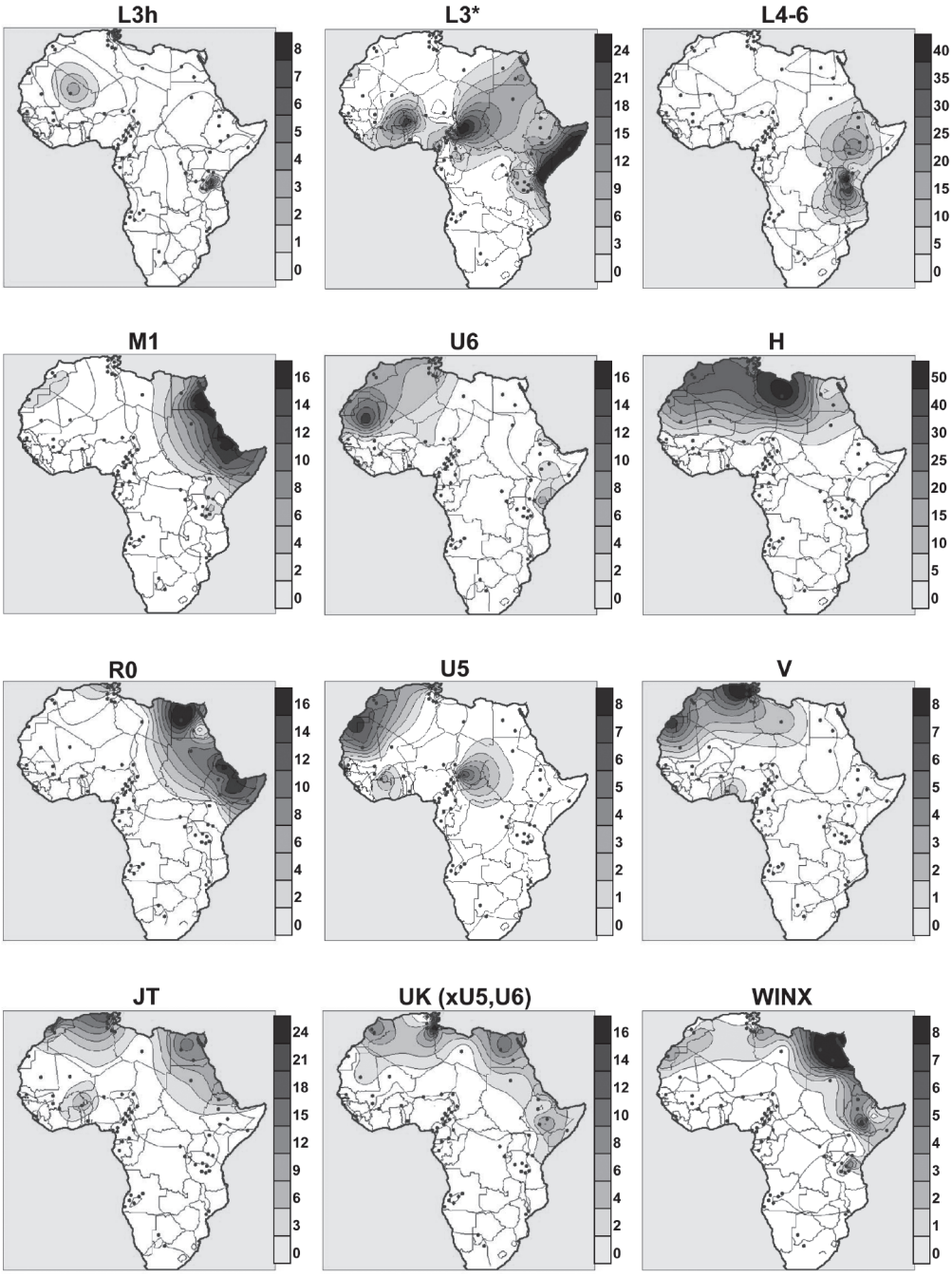

Fig. 3 - Frequency distribution maps for L3h, L3*, L4-6, M1, U6, H, R0, U5, V, JT, UK (xU5,U6), WINX haplogroups present in Africa. Maps were generated using the Surfer 8.02 software and redrawn by hand. Populations used to generate these frequency clines are shown in Appendix.

is strongly recommended when interpreting L3h sample assignment based solely on the HVS-I motif, without testing coding region diagnostic polymorphisms (*e.g.* Behar *et al.*, 2008). The currently described lineages within L3h are found to coalesce at approximately 70 kya (Behar *et al.*, 2008; Soares *et al.*, 2009) (Fig.1). At the present state-of-the-art, a panoply of L3* lineages yet to be unambiguously classified are thus included in the L3* paragroup. In the present review, we also pooled the L3c, L3j, L3i, L3k, L3x minor clades together with L3* (x L3b, d, e, f, h), given that the phylogeny resolution of early works here considered did not allow their classification.

Non-L haplogroups

Haplogroup M1

M1 lineages have an African supra-equatorial distribution being mainly present and more diverse in Northeastern and eastern Africa (Fig.3). Occasional occurrences are registered in West and Northwest Africa (Rando *et al.*, 1998: Rosa *et al.*, 2004; Cherni *et al.*, 2009; Ottoni *et al.*, 2009) (Fig.3). Nevertheless, their distribution is not restricted to Africa, being also relatively common in the Mediterranean and peaking in the Iberia Peninsula, reaching even the Basque country. M1 has also a well established presence in the Caucasus and the Middle East, ranging from the Arabian Peninsula to Anatolia and from the Levant to Iran. Central Asian studies have reported its presence as far as Tibet (for a detailed list of references on M1 distribution see Appendixes 1 and 2 in González *et al.*, 2007).

For many years, a hot debate centred on the M1 origin, which is the only M representative nowadays found in Africa. Controversy concerned its African versus non-African origin with two possible scenarios: either *i)* M1 lineages originated in East Africa and sporadically spread within the continent or *ii)* their origin can be traced back outside Africa, entering the continent in a "back-to-Africa" migration (Quintana-Murci *et al.*, 1999; Tambets *et al.*, 2000; Maca-Meyer *et al.*, 2001; Forster, 2004; Kivisild *et al.*, 2004). Latest works seem to favor the West Eurasian/Near Eastern origin of M1 molecular ancestors (Olivieri *et al.*, 2006; González *et al.*, 2007). M1 lineages trace back 35-40 kya (Olivieri *et al.*, 2006) (Fig.1), although González *et al.* (2007) found a much younger coalescence age for the whole clade (20-30kya), in any case a younger coalescence age than those of Asian-exclusive M lineages (Sun *et al.*, 2006). Further evidence to the Near Eastern origin and eastwards dispersals into Central Asia and westwards migrations into Africa (possibly via the Sinai Peninsula) is added by M1c basal presence in Jordanians and as far as Tibet (González *et al.*, 2007).

More recently, spotlights were placed on the timeframe at which a "back-to-Africa" migratory flow introduced M1 lineages into the continent. Again, the analysis of variation and intricate geographic distribution of M1a, M1b and M1c subclusters favoured an ancient arrival at about 30 kya (González *et al.*, 2007), as already suggested by HVS-I lineages (Quintana-Murci *et al.*, 1999; Richards *et al.*, 2003; Forster, 2004; Kivisild *et al.*, 2004). Gonzalez *et al.* (2007) proposes an M1 first expansion towards northwestern African areas rather northeastern, eventually reaching Iberia. Furthermore, there is an absolute lack of other Asian-specific clades within M in Africa, as would be expected in the case of a more recent arrival. This is also coherent with the close match of M1 lineages and Afro-Asiatic linguistic phylum, therefore enhancing the East African/Nile Valley origin of this language family supported by scholars (Ehret *et al.*, 2003).

Haplogroup U6

Haplogroup U6 is mostly frequent in Northwest Africans, being rather frequent in Algerian Berbers, Moroccans and Mauritanians (Rando *et al.*, 1998; Cherni *et al.*, 2009) (Fig.3), but it is also present in Eastern Africans (Watson *et al.*, 1997; Kivisild *et al.*, 2004) (Fig.3). This haplogroup is seen as the first return to Africa of ancient Caucasoid lineages (40-50 kya; Rando *et al.*, 1998; Maca-Meyer *et al.*, 2003; Olivieri *et al.*, 2006). From a West Eurasian/Near Eastern source the most parsimonious scenario is that of a

joint diffusion of M1 and U6 and arrival in North Africa in the Early Upper Paleolithic at about 30 kya, likely a human retreat to Africa forced by a harsh glacial period (González *et al.*, 2007).

The most representative of U6 clades, U6a, displays increasing frequency and diversity towards Northwest Africa, supporting the idea of an autochthonous prehistoric lineage at about 38 kya (Olivieri *et al.*, 2006). More recent local dispersal of U6b and U6c sub-clades also parallels the distribution of M1b and M1c1 (González *et al.*, 2007). One of the most frequent U6a lineages is believed to have started to expand ~11 kya and partially diffused to the Sahel (Watson *et al.*, 1997; Rando *et al.*, 1998; Brehm *et al.*, 2002; Cerný *et al.*, 2004; Rosa *et al.*, 2004; Coia *et al.*, 2005; Jackson *et al.*, 2005).

Haplogroup U5

The spatial and temporal overlap for the origin of haplogroups M1 and U6 (Olivieri *et al.*, 2006) is also true for U5 (Richards *et al.*, 2000; Achilli *et al.*, 2005) (Fig.3), which raises the possibility that their molecular ancestors lived in the same broad geographical area of Southwest Asia (possibly in separate regional enclaves). The main radiation of U5 took place in Europe, where it arrived in early Upper Palaeolithic and was likely among the first AMHs peopling the continent, most probably from Middle East/Caucasus region (~40-50 kya; Richards *et al.*, 2000; Olivieri *et al.*, 2006). An unexpected finding concerning the recent U5b1b branch linked the Scandinavian Saami to the North African Berbers and the sub-Saharan Fulani, not earlier than ~9 kya (Achilli *et al.*, 2005). Together with H1, H3 and V, also commonly found along the Mediterranean coast of Africa (Fig.3), these lineages are interpreted as post-LGM signatures (Achilli *et al.*, 2004, 2005). The Franco-Cantabrian refuge area is then seen as the source of late-glacial surviving lineages carried by the hunter-gatherers that later repopulated most of Europe, and also contributed to the mtDNA pool of North Africans by crossing the Strait of Gibraltar (Achilli *et al.*, 2005). Specific sub-clades of U5b were found at very low frequencies across sub-Saharan West Africa (e.g. Rando *et al.*, 1998; Rosa *et al.*, 2004; Cerný *et al.*, 2006) giving support to the hypothesis of hitch-hiking episodes of North Africans that crossed the Sahara, through the establishment of commercial networks.

Other N- and R-derived haplogroups

The presence of haplogroup X in West Eurasia dates back to pre-Holocene at about 30 kya (Reidla *et al.*, 2003), soon after it had diverged from superhaplogroup N as did its W, N1 (and I) sister clades (Fig.1). Nowadays, it is present at low frequencies among West Eurasian, North Africa and Near East, with specific sub-clades in Native Americans (Reidla *et al.*, 2003 and references therein). Clade X1 is largely restricted to Afro-Asiatic populations in North Africa (particularly Moroccans) and Ethiopians (Kivisild *et al.*, 2004), suggesting a diffusion along the Mediterranean and Red Seas (Reidla *et al.*, 2003). The coalescence age of this clade in North Africa is contemporary to M1 and U6 Paleolithic mtDNA lineages (Reidla *et al.*, 2003).

The African maternal pool of those populations inhabiting the Mediterranean coast became enriched in Eurasian lineages, thanks to several migratory episodes. The dispersal timeframe of H lineages (other than H1 and H3) in Egypt is comparable to that of Jordanian H mtDNAs (respectively 21 kya and 23 kya; Roostalu *et al.*, 2007; Cherni *et al.*, 2009). This drives us to believe in temporarily coincident events in North Africa and the Middle East, followed by southward migrations due to the cold climate of Eurasian LGM 20 to 14 kya. X2 lineages probably arose and dispersed around or after the glacial period not earlier than 25 kya, which is consistent with their geographic range across all Europe, the Near East (where it is more diverse), and North Africa, and wider than that of its X1 sister clade (Richards *et al.*, 2000; Reidla *et al.*, 2003).

Haplogroups J, T1 and R0 have a clear Middle/Near Eastern origin stated in their age estimates and declining frequency towards the southern Caucasus, Near East, Europe and North and East Africa (Rando *et al.*, 1998; Krings *et al.*, 1999; Macaulay *et al.*, 1999; Richards *et al.*, 2000;

Maca-Meyer *et al.*, 2001; Kivisild *et al.*, 2004; Kujanová *et al.*, 2009) (Fig.3). Their spread to Europe and North Africa is associated with the Neolithic movements of farming-herding societies 10 to 8 kya. There is nevertheless controversy about the demic input of Near Eastern Neolithic on the Northwest African autochthonous Capsian Neolithic culture (Newman, 1995) These maternal lineages may have also diffused to North Africa in more recent times, perhaps as lineages of Phoenician traders.

N1 is a minor mtDNA haplogroup that has been observed at marginal frequencies in European, Near Eastern, Indian and East African populations (Richards *et al.*, 2000; Kivisild *et al.*, 2004), mainly in Semitic speakers. Although with much older coalescences in the Near East and Southwest Asia (Richards *et al.*, 2000), N1, U (non-U5 or U6) and W lineages may have been imported relatively recently, with the expansion of Semitic languages, at least in the Ethiopian pool (Kivisild *et al.*, 2004).

Regional population patterns of African mtDNA variation

Despite the many years of research and efforts to develop more accurate models and statistical tools for interpreting the phylogenetic and phylogeographic pattern of mtDNA genetic variation, one should not expect it to directly reflect demographic events from the initial occupation to the recent historical episodes. The earlier pattern of mtDNA distribution on the African continent was most likely altered by countless processes over the last couple of hundred thousand years, among the best known and which can be traced on a genetic basis are the post-LGAM and agricultural demographic growth. On the other hand, episodes in recent history are not readily detectable, and are often not the main focus of attention of population geneticists. Furthermore, the evolutionary relationships based on uniparentally transmitted polymorphisms are primarily concerned with the history of genes and not of populations. Any suggestions on a population-based phylogeographic approach are mere hypotheses built on genetic evidence, which caresses support from non-genetic data. The peopling and migratory processes in Africa remain poorly understood and are further complicated by the unfortunately weak preservation of archaeological and anthropological elements, given the chemical and physical properties of soil. Once conscious of the particularities and limitations of the genetic systems, this should not hinder us from attempting to interpret the extant variation. Last, but not least, the complex African ethnolinguistic context must be taken into consideration, given where societies are deeply structured by cultural constraints and patterns of admixture (endogamy or patrilocal exogamy, for instance), which strongly influence the mtDNA inheritance pattern. Exploring each of these themes could be the topic for a handful of extensive reviews. Therefore, we here chose to stress only particular populational groups, as follows.

The mtDNA pool of the Khoisan people shows over 60% of L0d and L0k lineages (Watson *et al.*, 1997; Chen *et al.*, 2000; Salas *et al.*, 2002; Gonder *et al.*, 2006; Behar *et al.*, 2008), the deepest known genetic clades in modern humans. Therefore, these are considered by many as unique relics of the Middle Stone Age hunter-gatherer genetic pool, and the initial source of L0d'k in non-Khoisan. Although the co-evolution of L0-L6 lineages in a single East/Southeast African population has been proposed (Maca-Meyer *et al.*, 2001; Forster, 2004; Kivisild *et al.*, 2004; Macaulay *et al.*, 2005; Gonder *et al.*, 2006), the hypothesis of an L0-L1'6 split, representing both a phylogenetic and populational split into two small populations in South and East Africa, was recently put forward by Behar *et al.* (2008). It seems rather unlikely that only basal L0d and L0k became enriched by drift in the Khoisan while following extinction in non-Khoisans. Based on their present-day particular mtDNA pool, these authors support an extensive maternal genetic structure in the early evolution as a result of an ancient populational (and phylogenetic) split at 210-140 kya (Behar *et al.*, 2008), and are not consistent with a homogenous distribution

of modern humans throughout sub-Saharan Africa as proposed before (Watson *et al.*, 1997; Forster, 2004). The introgression of L0abf lineages on the geographic range of L1'6 could then be explained by a dispersal event *circa* 144 kya, likely due to more favorable environmental conditions in eastern Africa (Mellars, 2006). In fact, evidence of modern human behavior, namely of a marine diet is documented in East Africa ca. 125 kya (Stringer, 2000). The divergence of the maternal pool of Khoisan ancestors from other humans probably occurred 140-90 kya, followed by independent evolution of L0d'k in the early southern African population. The same authors also believed that only much later, around 40 kya, introgression of additional lineages occurred, a process further accelerated by the expansion of Bantu-speaking people (Behar *et al.*, 2008). On the PCA displayed in Fig.4, South African Khoisan are driven by their proportion of L0dk. Nevertheless, these are placed nearby East, Central and South African Bantu-speakers and Central African Niger-Congo Adamawa and Afro-Asiatic-Chadic speakers, given that L0af and L0dk are here shown as nearly overlapping vectors, therefore forcing their position on the plot. On the other hand, the Tanzanian click-speakers are integrated on a cluster under a geographic reasoning, harboring Khoisan, Nilo-Saharan Nilotic and Afro-Asiatic Cushitic people, all inhabitants of Tanzania. These last are under the influence of their L4-L6 mtDNAs, and corroborate the existence of gene flow with the surrounding populations.

The proposed pattern of sequestration in independent small communities offers a very plausible theory that only L3-derived M and N clades left Africa and gave rise to all non-African mtDNA variation (Behar *et al.*, 2008 and references therein). This is also a likely scenario for the ancestors of contemporary Central African Pygmies. To the present-day, Pygmy hunter-gatherers refuse to adopt the agricultural lifestyle and remain geographically restricted to western and eastern groups in Central Africa Cameroon, Gabon and the Central African Republic and Democratic Republic of Congo. Their genetic pool, enriched in L1c and L5 mtDNA lineages with roughly 95 and 130 ky of age, renders them the title of "living-fossils", when compared to other sub-Saharans. According to data in Quintana-Murci *et al.* (2008), the Central African region was colonized with an initial L1c-rich ancestral population (alternatively gave rise to L1c *in situ*), which was on average 25 ky older than L0, L2 and L3 types that were found to later enrich their pool. The ancestral population probably split earlier than 70 kya and, in isolation, diverged into distinctive clades, many of which are today frequently found among the Bantu agriculturalists and the Western Pygmies (Batini *et al.*, 2007; Quintana-Murci *et al.*, 2008). The distinctive physical and cultural characteristics of Pygmies are thought to reflect the long-term isolation and adaptation to the rain-forest. The isolation period was interrupted by gene flow at about 40 kya, when both populations were probably smaller. Furthermore, Pygmies show small population sizes and strong genetic drift while food-producers have high levels of haplogroup diversity though more genetically homogenous among them. Quintana-Murci and colleagues (2008) reported a higher level of asymmetric gene flow between Western Pygmies and Bantu-speakers, while Eastern Pygmies showed a haplogroup pattern more alike to that of the eastern African population. This is corroborated by their relative position in the plot of Figure 4. Nevertheless, the Bantu-to-Pygmy gene flow probably did not happen before the expressive growth of Bantu people 4 kya. As for the Y-chromosome (Pereira *et al.*, 2002; Beleza *et al.*, 2005; Wood *et al.*, 2005), mtDNA data reflect the strong influence of socio-cultural factors, namely the unidirectional marriages between hunter-gatherer females and food-producer males (Destro-Bisol *et al.*, 2004; Quintana-Murci *et al.*, 2008).

In the millennia following the early spread, genetic variation accumulated in small independent communities (Forster, 2004; Behar *et al.*, 2008). Records of geometric notches in bone dating nearly 70 kya were found in South Africa (Henshilwood *et al.*, 2002), which temporally overlaps with signs of a fragmented environment

(Lahr & Foley, 1998; Mellars, 2006). Major demographic and range expansions at about 60 kya are evident in the starlike phylogenies and wide distribution of the main L2 and L3 clades that have repopulated Africa (Watson *et al.*, 1997; Gonder *et al.*, 2006; Behar *et al.*, 2008).

The archaeological findings support a permanent occupation of West Africa by modern humans from 30-40 kya onwards (Mercader & Marti, 2003) or possibly even earlier (Phillipson, 1993; Newman, 1995; Foley & Lahr, 1997; Cornelissen, 2002). Such evidence is concordant with dated ancestral mtDNA types that seem to have given rise to autochthonous branches, with approximate coalescence time at 30-40 kya, namely L1b, L2b, L3b and L3d (Watson *et al.*, 1997; Rando *et al.*, 1998; Salas *et al.*, 2002; Rosa *et al.*, 2004). Subsequent climatic oscillations between 40-12 kya had a major impact on the modern sub-Saharan mtDNA phylogeographic sub-structuring. The cycles of expansion and retraction of the equatorial forest (Adams & Faure, 1997; Foley & Lahr, 1997; Cornelissen, 2002) was an ecological scenario able to reduce the genetic diversity of AMHs, and generate a fragmented pattern of population distribution, ultimately evolving into regional-specific clusters. Therefore, the first inhabitants of West Africa were probably dispersed in small and isolated hunter-gatherer groups, and their genetic pool was unlikely to have been uniform.

The climatic oscillations at the LGAM 23-15 to 9 kya, promoted the reduction of woodlands and savannas to small vegetation pockets south of the Sahelian strip (Adams & Faure, 1997). These may have acted as refuge areas conserving the genetic legacy, and from where modern humans later spread. We observe today an overall similarity in the West African mtDNA pool, with a few lineages exhibiting signs of expansion and founder types shared by different ethnic groups in West Africans. This is evidenced in Figure 4 by the main influence of haplogroups L2a-d, L3b and L3d and likely testifies for a common basis and co-evolution of the genetic diversity that has emerged from the refugia. . Nevertheless, a few populational units exhibit similarities of their maternal pool with Central African groups. In fact, and in what concerns the great ethnic diversity of West and Central Africans, neither linguistics nor geography are good predictors of population affinities. The return to moister and warmer conditions culminated in the Sahara's wet phase ~9kya (Aumassip *et al.*, 1994), which is believed to have supported populational growth and massive displacement of people, reaching uninhabited areas and allowing contacts and gene flow between previously isolated people (Camps, 1974; Hassan, 1978; Dutour *et al.*, 1988; Clark *et al.*, 1994). As stated before, particular clades of L2a, L2c and L1b are believed to have expanded east- and westwards along the Sahel corridor, after the LGAM (Salas *et al.*, 2002; Rosa *et al.*, 2004). The range expansion of L3e2a mtDNAs to West Africa (Salas *et al.*, 2002; Rosa *et al.*, 2004) suggests these as successful hitchhikers of Saharan movements during the Holocene Great Wet Phase and subsequent Wet Phase (Muzzolini *et al.*, 1993; Bandelt *et al.*, 2001).

Less frequent clades in West Africa, both L and non-L types, point towards Central, East and even North African gene flow occurring at different timeframes. As mentioned above, the U5b1b lineages are post-glacial signatures of lineages that have crossed the strait of Gibraltar towards Northwest Africa, and further develop into West and Central African local clusters (Rando *et al.*, 1998; Rosa *et al.*, 2004; Coia *et al.*, 2005; Cerný *et al.*, 2006; Ely *et al.*, 2006). Its non-random sub-Saharan spread was likely mediated by the Berbers or Berber- related people, such as the Fulani (Rosa *et al.*, 2004; Achilli *et al.*, 2005). The Fulani have a proposed ancient origin on the northerly mountain massifs of the Central Sahara (Dupuy *et al.*, 1999). These are nowadays mostly nomadic pastoral communities whose gene flow with local agricultural populations is limited to a few exceptions of women paid as a tribute to freely circulate in other territories (Almada, 1964). As displayed in Figure 4, the L1b lineages are typically frequent among the Fulani and Serer and Wolof people with whom close relationships are maintained, driving their clustering (except for North Cameroon Fulani).

Furthermore, a few Fulani-exclusive haplotypes and exact mtDNA matches on several haplogroups among Fulani individuals inhabiting a broad geographic area in West-Central Africa (Watson *et al.,* 1997; Destro-Bisol *et al.*, 2004; Rosa *et al.*, 2004; Cerny *et al.*, 2006). This most likely tells of their largely common ancestry, with a lower differentiation of the maternal pool among nomadic Fulani kept by ongoing gene flow, while the settled communities accumulated differences, by gene flow with their geographic neighbors (Rosa *et al.*, 2004; Cerny *et al.*, 2006).

The size of African human populations probably slowly increased in the Late Pleistocene-Holocene, limited by the available fauna. At about 14 kya, when animals were driven to near extinction by hunting, the impetus for people to adopt cultivation as a subsistence strategy was triggered, and foraging was progressively abandoned (Cohen, 1989). By 6 kya, centers in the Sahel temperate sub-tropical areas were cultivating local crops (Atherton, 1972; Calvocoressi & David, 1979; Clark *et al.*, 1994). Better nutrition supported a larger number of people (Diamond & Bellwood, 2003), but one cannot easily distinguish which subsets of the extant variation were generated by the post-LGAM return to more beneficial conditions or by the subsequent shift to agriculture. Among these are the L0a1and L3h lineages with West African MRCA ages of 6.5 ky, although with a certainly prior defined diversity (Rosa *et al.,* 2004). More likely, the populational growth happened in a continuous timescale since the climatic return to more favorable conditions and became more accentuated with the introduction of agriculture and iron smelting techniques, obscuring earlier diversity and many differences accumulated during the isolation periods. Although a much older clade, L3e4 is thought to signal expansion events related to food production and iron-smelting in Atlantic West Africa (Bandelt *et al.,* 2001; Rosa *et al.,* 2004).

Mande people inhabiting West Africa are physically and culturally descendent from the Mali Empire, which controlled the trans-Saharan trade from the Middle East to West Africa. Moreover, and according to both genetic and archaeological data, the ancestors of these indigenous people may relate to the people who instigated farming expertise in West Africa (Cavalli-Sforza *et al.*, 1994; Fage, 1995; Rosa *et al.*, 2004). The food-producing economy has therefore supported a substantial populational increase and expansions in the Niger/Mali/Burkina-Faso region (Cavalli-Sforza *et al.*, 1994). On the account of several L2 and L3 particular branches, West African Mandenka form a sub-cluster within the shown West African cluster.

The Bantu migrations, one of the most important recent demographic upheavals in African history, supposedly started at about 3kya from a central African source in the vicinity of the Cross River Valley (Phillipson, 1993). The movement is associated with the transition from a hunter-gathering to agricultural lifestyle and the advent of iron-smelting (Newman, 1995; Phillipson, 1993). Conditions for population growth were created, with the consequent spatial expansion following two main east- and westwards spread routes to the south. Although not forming an individual cluster and with few exceptions, the relative positions of Bantu-people on the PCA on the main responsibility of L0af, L1c and L3e clades, despite their Central, East or South African location, corroborates their origin and southwards routes (Fig.4; vectors not shown). There is genetic evidence from both the mtDNA and Y chromosome systems that testify the strong impact of the Bantu migrations on the gene pool of sub-Saharan Africa, almost erasing the pre-existent one. The genetic evidence is in agreement with archaeological and linguistic findings suggesting a gradual spread of Bantu-speakers and strong interactions (and gene flow) between different lines of Bantu-speaker descent and other populations *en route* (Ehret, 2001; Tishkoff *et al.,* 2007; Castrì *et al.,* 2009). This is seen as the main mechanism underlying the spread of the so-called "Bantu-markers" mentioned above.

North Africans are genetically quite different from sub-Saharans, which is reflected in their mtDNA haplogroup composition. The sub-Saharan portion represents less than half of their pool although all the major L0, L1, L2 and L3 branches are present (Rando *et al.,* 1998; Salas *et al.,* 2002;

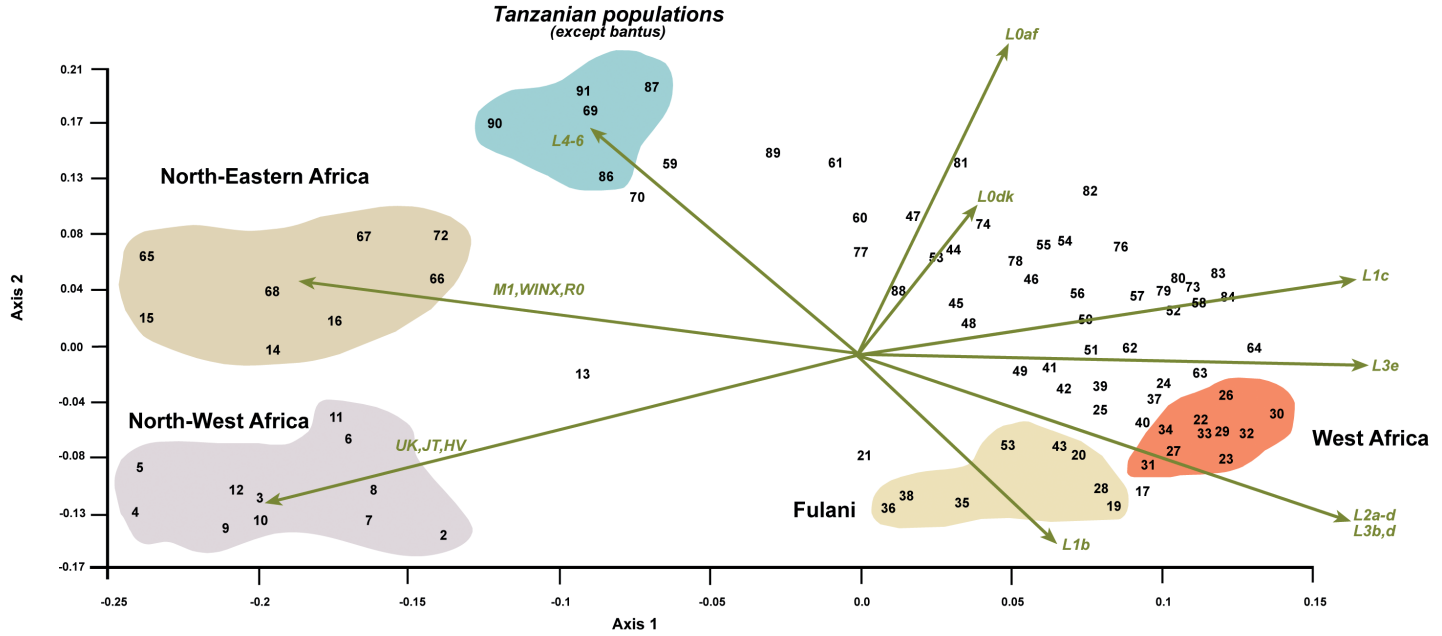

Fig. 4 - Principal Component Analysis of African populations based on mtDNA haplogroup relative frequencies. The analysis gives us the topological distribution of the maternal pool of African populations listed in Appendix. Given their "mixed" linguistic affinities and in order to improve the resolution of tight clusters, Mauritanians (1), Senegalese (18), Nairobians (71), Mozambicans (75) and Angolans (85) were excluded from the analysis. Furthermore, paragroups L1, L2* and L3* were not considered, as these may include different yet unclassified lineages. The two principal components displayed account for 28.0% and 15.9% of the total variation, respectively. With a few exceptions, the first component establishes a geographical clear-cut between Northwest/Northeast/East Africa and West/Central/South Africa. The second component separates Northeast/East/Central/South Africa from Northwest/West Africa.*

González *et al.,* 2006; Cherni *et al.,* 2009; Ottoni *et al.,* 2009). The remaining pool is of West Eurasian origin, most haplogroups affiliated with a clear Near Eastern origin. Climatic changes since the Late Pleistocene to the present day allowed intense cultural developments and permitted cross-migrations between Africa and Asia. Among the first to arrive in the south Mediterranean basin, as a Near Eastern backflow, are M1 and U6 (and perhaps X1) at about 30 kya (Reidla *et al.*, 2003; Olivieri *et al.,* 2006; González *et al.*, 2007). Among the later back migrants worth mentioning are H and X2 as post-LGM migrants, and J, T1 and R0 as Neolithic influences (Richards *et al.*, 2000; Reidla *et al.*, 2003). The diffusion of less frequent N-descendant clades such as N1, W and I is less understood, but it may trace back to episode(s) later than their Near Eastern coalescence nearly 25 kya to recent historical times. The genetic mtDNA profile of Semitic and Cushitic Afro-Asiatic speakers, inhabitants of Northwest Africa here considered, are separate from other population clusters mainly on account of several Eurasian mtDNA haplogroups, namely JT, H, V, UK and R0 (see Fig.4). Other Afro-Asiatic speakers in Northeast/East Africa, to whom Nilo-Saharan Nubians are in close proximity, cluster on the responsibility of M1 and minor W, I, N, X haplogroups (Fig.4, vectors not shown).

The Middle East was not the only source for Eurasian haplotypes of North Africans. More data has been recently added to support the hypothesis of an ancient contact between the Iberian Peninsula and North West Africa via the Gibraltar Strait, including the analysis of complete mtDNA sequences, at least in what concerns post-LGM sub-haplogroups H1, H3, V and U5b1b (Torroni *et al.*, 2001a; González *et al.*, 2003; Achilli *et al.*, 2004, 2005; Loogväli *et al.*, 2004; Pereira *et al.*, 2005; Olivieri *et al.*, 2006; Cherni *et al.*, 2009; Ennafaa *et al.*, 2009).

The trans-Saharan spread of North African mtDNA legacy is not likely to be a product of recent gene flow, since a random assortment of Northwest African mtDNAs would have also been carried by the migrants. We have to bear in mind that, even with well documented commercial networks, the Sahara desert seems to be an important barrier to southward gene flow of North Africa Eurasiatic haplogroups (González *et al.*, 2006).

(Large-scale) Out-of-Africa compulsory gene flow in recent times

The current distribution range of African mtDNA lineages is far broader than the African continent. Long-distance gene flow mediated by the Atlantic slave trade since the 16th century is worth mentioning in this review. Brazilians harbor the most important reservoir of African maternal lineages outside of Africa. Early description of the genetic landscape of Brazilians with sub-Saharan ancestry confirms the historical evidence, with L1c and L3e lineages summing up to nearly half of the African share (Alves-Silva *et al.*, 2000). Later studies on Afro-Americans residing in the American continent report 65% of mtDNA types in South America as having a Central African origin, 41% and 59% of Central Americans tracing progeny to West Central Africa and West Africa respectively, while North American ancestors are estimated as being 28% West-Central Africans and 72% West Africans (Salas *et al.*, 2004). These results corroborate the historical record of these regions (Thomas, 1998). The origin of Afro-Americans in U.S.A. is associated with West African (>55%) and West-Central/Southwest African (<45%) mothers, also in close proximity to historical data (McMillin, 2004).

Recent results on admixture analysis suggest that Africans brought to Brazil as slaves were originally from two geographical regions: *i)* 69% of the maternal pool of Black Brazilians in Rio de Janeiro is attributed to West-Central and Southeast Africa, close to two former Portuguese colonies (Angola and Mozambique) and *ii)* 82% of mtDNA lineages in Porto Alegre are found in West Africa, in the northern portion of the Gulf of Guinea (Hünemeier *et al.*, 2007). Such detailed analysis is possible given the clear mtDNA haplogroup structure which allows the discrimination of geographic/linguistic origins. Once again,

genetic records are in agreement with historical data (Klein, 2002). Moreover, these authors state the absence of major geographic gender specific differences in the Atlantic slave trade (Hünemeier *et al.*, 2007), contradicting what was suggested before by Silva *et al.*, 2006). Though indirectly, African lineages found outside of Africa may bring additional information to aid the reconstruction of evolutionary demographic events mediated by women in African history and pre-history, despite hitherto non-sampled regions.

References

Achilli A., Rengo C., Battaglia V., Pala M., Olivieri A., Fornarino S., Magri C., Scozzari R., Babudri N., Santachiara-Benerecetti A.S., Bandelt H.J., Semino O. & Torroni A. 2005. Saami and Berbers - an unexpected mitochondrial DNA link. *Am. J. Hum. Genet.,* 76: 883-886.

Achilli A., Rengo C., Magri C., Battaglia V., Olivieri A., Scozzari R., Cruciani F., Zeviani M., Briem E., Carelli V., Moral P., Dugoujon J.M., Roostalu U., Loogväli E.L., Kivisild T., Bandelt H.J., Richards M., Villems R., Santachiara-Benerecetti A.S., Semino O. & Torroni A. 2004. The Molecular Dissection of mtDNA Haplogroup H Confirms That the Franco-Cantabrian Glacial Refuge Was a Major Source for the European Gene Pool. *Am. J. Hum. Genet.,* 75: 910-918.

Adams J.M. & Faure H., 1997. Review and Atlas of Palaeovegetation: Preliminary land ecosystem maps of the world since the Last Glacial Maximum. Oak Ridge National Laboratory, TN, USA. http://www.esd.ornl.gov/projects/qen/adams1.html

Almada A.A., 1964. Tratado breve dos rios da Guiné e de Cabo Verde. Editorial LIAM, Lisboa.

Alves-Silva J., da Silva Santos M., Guimarães P.E., Ferreira A.C., Bandelt H.J., Pena S.D. & Prado V.F. 2000. The ancestry of Brazilian mtDNA lineages. *Am. J. Hum. Genet.,* 67: 444-461.

Anderson S., Bankier A.T., Barrell B.G., de Bruijn M.H., Coulson A.R., Drouin J., Eperon I.C., Nierlich D.P., Roe B.A., Sanger F., Schreier P.H., Smith A.J., Staden R. & Young I.G. 1981. Sequence and organization of the human mitochondrial genome. *Nature,* 290: 457-465.

Andrews R.M., Kubacka I., Chinnery P.F., Lightowlers R.N., Turnbull D.M. & Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.,* 23: 147.

Arredi B., Poloni E.S., Paracchini S., Zerjal T., Fathallah D.M., Makrelouf M., Pascali V.L., Novelletto A. & Tyler-Smith C. 2004. A predominantly Neolithic origin for Y-chromosomal DNA variation in North Africa. *Am.J.Hum. Genet.,* 75: 338-345.

Atherton J.H. 1972. Excavations at Kamabai and Yagala Rock Shelters, Sierra Leone. *West Afr. J. Archaeol.,* 2: 39-74.

Atkinson Q.D., Gray R.D. & Drummond A.J. 2009. Bayesian coalescent inference of major human mitochondrial DNA haplogroup expansions in Africa. *Proc. Biol. Sci.,* 276: 367-373.

Aumassip G., Ferhat N., Heddouche A. & Vernet R., 1994. *Le milieu saharien aux temps prehistoriques, Milieux, hommes et techniques du Sahara prehistorique. Problémes actuels*. L'Harmattan, Paris, pp. 9-29.

Avise J.C., 2000. *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts.

Bandelt H.J., Alves-Silva J., Guimarães P.E., Santos M.S., Brehm A., Pereira L., Coppa A., Larruga J.M., Rengo C., Scozzari R., Torroni A., Prata M.J., Amorim A., Prado V.F. & Pena S.D. 2001. Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade. *Ann. Hum, Genet.,* 65: 549-563.

Bandelt H.J. & Forster P. 1997. The myth of bumpy hunter-gatherer mismatch distributions. *Am. J. Hum. Genet.,* 61: 980-983.

Bandelt H.J., Forster P. & Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.,* 16: 37-48.

Bandelt H.J., Forster P., Sykes B.C. & Richards M.B. 1995. Mitochondrial portraits of human populations using median networks. *Genetics,* 141: 743-753.

Bandelt H.J., Kong Q.P., Richards M., Villems R. & Macaulay V., 2006. *Estimation of mutation rates and*

coalescence times: some caveats, Human Mitochondrial DNA and the Evolution of Homo sapiens, pp. 149-179. Springer-Verlag, Berlin, Heidelberg.

Bandelt H.J., Macaulay V. & Richards M. 2000. Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. *Mol. Phylogenet. Evol.,* 16: 8-28.

Batini C., Coia V., Battaggia C., Rocha J., Pilkington M.M., Spedini G., Comas D., stro-Bisol G. & Calafell F. 2007. Phylogeography of the human mitochondrial L1c haplogroup: genetic signatures of the prehistory of Central Africa. *Mol. Phylogenet. Evol.,* 43: 635-644.

Behar D.M., Villems R., Soodyall H., Blue-Smith J., Pereira L., Metspalu E., Scozzari R., Makkan H., Tzur S., Comas D., Bertranpetit J., Quintana-Murci L., Tyler-Smith C., Wells R.S. & Rosset S. 2008. The dawn of human matrilineal diversity. *Am. J. Hum. Genet.,* 82: 1130-1140.

Beleza S., Gusmao L., Amorim A., Carracedo A. & Salas A. 2005. The genetic legacy of western Bantu migrations. *Hum. Genet.,* 117: 366-375.

Brandon M.C., Ruiz-Pesini E., Mishmar D., Procaccio V., Lott M.T., Nguyen K.C., Spolim S., Patil U., Baldi P. & Wallace D.C. 2009. MITOMASTER: a bioinformatics tool for the analysis of mitochondrial DNA sequences. *Hum. Mutat.,* 30: 1-6.

Brandstätter A., Peterson C.T., Irwin J.A., Mpoke S., Koech D.K., Parson W. & Parsons T.J. 2004. Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database. *Int. J. Legal. Med.,* 118: 294-306.

Brehm A., Pereira L., Bandelt H.J., Prata M.J. & Amorim A. 2002. Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade. *Ann. Hum. Genet.,* 66: 49-60.

Brown W.M. 1980. Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proc. Natl. Acad. Sci. U.S.A.,* 77: 3605-3609.

Calvocoressi D. & David N. 1979. A new survey of radiocarbon and thermoluminescence dates for West Africa. *J. Afr. Hist.,* 20: 1-29.

Camps G., 1974. Les Civilisations Préhistorique de l'Áfrique du Nord et du Sahara. Paris, Doin.

Cann R.L., Stoneking M. & Wilson A.C. 1987. Mitochondrial DNA and human evolution. *Nature,* 325: 31-36.

Castri L., Tofanelli S., Garagnani P., Bini C., Fosella X., Pelotti S., Paoli G., Pettener D. & Luiselli D. 2009. mtDNA variability in two Bantu-speaking populations (Shona and Hutu) from Eastern Africa: implications for peopling and migration patterns in sub-Saharan Africa. *Am. J. Phys. Anthropol.*,140: 302-311.

Cavalli-Sforza L.L., Menozzi P. & Piazza A., 1994. *The History and Geography of Human Genes*. Princeton University Press, Princeton.

Cerný V., Fernandes V., Costa M.D., Hajek M., Mulligan C.J. & Pereira L. 2009. Migration of Chadic speaking pastoralists within Africa based on population structure of Chad Basin and phylogeography of mitochondrial L3f haplogroup. *BMC Evol. Biol.,* 9: 63.

Cerný V., Hajek M., Bromova M., Cmejla R., Diallo I. & Brdicka R. 2006. MtDNA of Fulani nomads and their genetic relationships to neighboring sedentary populations. *Hum. Biol.,* 78: 9-27.

Cerný V., Hajek M., Cmejla R., Bruzek J. & Brdicka R. 2004. mtDNA sequences of Chadic-speaking populations from northern Cameroon suggest their affinities with eastern Africa. *Ann. Hum. Biol.,* 31: 554-569.

Cerný V., Salas A., Hajek M., Zaloudkova M. & Brdicka R. 2007. A bidirectional corridor in the Sahel-Sudan belt and the distinctive features of the Chad Basin populations: a history revealed by the mitochondrial DNA genome. *Ann. Hum. Genet.,* 71: 433-452.

Chen Y.S., Olckers A., Schurr T.G., Kogelnik A.M., Huoponen K. & Wallace D.C. 2000. mtDNA variation in the South African Kung and Khwe-and their genetic relationships to other African populations. *Am. J. Hum. Genet.,* 66: 1362-1383.

Cherni L., Fernandes V., Pereira J.B., Costa M.D., Goios A., Frigi S., Yacoubi-Loueslati B., Amor M.B., Slama A., Amorim A., El Gaaied A.B. & Pereira L. 2009. Post-last glacial maximum expansion from Iberia to North Africa revealed

by fine characterization of mtDNA H haplogroup in Tunisia. *Am. J. Phys. Anthropol.,* 139: 253-260.

Clark J.D., De Laet S.J., co-edited by Dani A.H., Lorenzo J.L. & Nunoo R.B., 1994. *Africa: From the appearance of Homo sapiens sapiens to the beginnings of food production, Volume I - Prehistory and the Beginnings of Civilization,* pp. 191-206. Routledge, New York.

Coelho M., Sequeira F., Luiselli D., Beleza S. & Rocha J. 2009. On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. *BMC Evol. Biol.,* 9: 80.

Cohen M.N., 1989. *Health and Rise of Civilization*. Yale University Press, New Haven & London.

Coia V., Destro-Bisol G., Verginelli F., Battaggia C., Boschi I., Cruciani F., Spedini G., Comas D. & Calafell F. 2005. Brief communication: mtDNA variation in North Cameroon: lack of Asian lineages and implications for back migration from Asia to sub-Saharan Africa. *Am. J. Phys. Anthropol.,* 128: 678-681.

Cornelissen E. 2002. Human responses to changing environments in Central Africa between 40,000 and 12,000 BP. *World Prehist.,* 16: 197-235.

Day M.H., Stringer C.B. & De Lumley M.A., 1982. *A reconsideration of the Omo-Kibish remains and the erectus-sapiens transition, L' Homo erectus et la place de l'homme de Tautavel parmi les hominides fossiles,* pp. 814-846. Centre National de la Recherche Scientifique, Nice.

Denaro M., Blanc H., Johnson M.J., Chen K.H., Wilmsen E., Cavalli-Sforza L.L. & Wallace D.C. 1981. Ethnic variation in Hpa 1 endonuclease cleavage patterns of human mitochondrial DNA. *Proc. Natl. Acad. Sci. U.S.A.,* 78: 5768-5772.

Destro-Bisol G., Donati F., Coia V., Boschi I., Verginelli F., Caglia A., Tofanelli S., Spedini G. & Capelli C. 2004. Variation of female and male lineages in sub-Saharan populations: the importance of sociocultural factors. *Mol. Biol. Evol.,* 21: 1673-1682.

Diamond J. & Bellwood P. 2003. Farmers and their languages: the first expansions. *Science,* 300: 597-603.

Dupuy C., Botte R., Boutrais J. & Schmitz J. 1999. *Les apports de l'archeologie et de l'ethnologie a la connaisance de l'histoire ancienne des Peuls, Figure Peules*, pp. 53-72. Karthala, Paris.

Dutour O., Vernet R., Aumassip G. & et al., 1988. *Le peuplement préhistorique du Sahara, Milieux, hommes et techniques du Sahara préhistorique. Problémes actuels*, pp. 39-52. L'Harmattan, Paris.

Ehret C., Renfrew C., McMahon A. & Trask L. 2003. Language family expansions: broadening our understandings of cause from an African perspective. In P. Bellwood & C. Renfrew (eds), *Examining the farming/language dispersal hypothesis,* pp. 163-176. McDonald Institute for Archaeological Research, Cambridge.

Ehret C. 2001. Bantu expansions: re-envisioning a central problem of early African history. *Int. J. African Hist. Studies,* 34: 5-41.

Ely B., Wilson J.L., Jackson F. & Jackson B.A. 2006. African-American mitochondrial DNAs often match mtDNAs found in multiple African ethnic groups. *BMC. Biol.,* 4: 34.

Ennafaa H., Cabrera V.M., Abu-Amero K.K., Gonzalez A.M., Amor M.B., Bouhaha R., Dzimiri N., Elgaaied A.B. & Larruga J.M. 2009. Mitochondrial DNA haplogroup H structure in North Africa. *BMC Genet.,* 10: 8.

Fage J., 1995. *A history of Africa*. Routledge, London.

Finnilä S., Hassinen I.E. & Majamaa K. 2001. Phylogenetic analysis of mitochondrial DNA in patients with an occipital stroke. Evaluation of mutations by using sequence data on the entire coding region. *Mutat. Res.,* 458: 31-39.

Foley R. 1998. The context of human genetic evolution. *Genome Res.,* 8: 339-347.

Foley R.A. & Lahr M.M. 1997. Modern technologies and the evolution of modern humans. *Cambridge Archeol. J.,* 7: 3-36.

Forster P. 2004. Ice Ages and the mitochondrial DNA chronology of human dispersals: a review. *Philos. Trans. R. Soc. Lond. B Biol. Sci.,* 359: 255-264.

Forster P., Harding R., Torroni A. & Bandelt H.J. 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. *Am. J. Hum. Genet.,* 59: 935-945.

Forster P., Röhl A., Lunnemann P., Brinkmann C., Zerjal T., Tyler-Smith C. & Brinkmann B. 2000. A short tandem repeat-based phylogeny for the human Y chromosome. *Am. J. Hum. Genet.,* 67: 182-196.

Forster P., Torroni A., Renfrew C. & Röhl A. 2001. Phylogenetic star contraction applied to asian and Papuan mtDNA evolution. *Mol. Biol. Evol.,* 18: 1864-1881.

Gonder M.K., Mortensen H.M., Reed F.A., de S.A. & Tishkoff S.A. 2006. Whole mtDNA Genome Sequence Analysis of Ancient African Lineages. *Mol. Biol. Evol.,* 24: 757-768.

González A.M., Brehm A., Perez J.A., Maca-Meyer N., Flores C. & Cabrera V.M. 2003. Mitochondrial DNA affinities at the Atlantic fringe of Europe. *Am. J. Phys. Anthropol.,* 120: 391-404.

González A.M., Cabrera V.M., Larruga J.M., Tounkara A., Noumsi G., Thomas B.N. & Moulds J.M. 2006. Mitochondrial DNA variation in Mauritania and Mali and their genetic relationship to other Western Africa populations. *Ann. Hum. Genet.,* 70: 631-657.

González A.M., Larruga J.M., Abu-Amero K.K., Shi Y., Pestano J. & Cabrera V.M. 2007. Mitochondrial lineage M1 traces an early human backflow to Africa. *BMC Genomics*, 8: 223.

Grun R., Beaumont P.B. & Stringer C.B. 1990. ESR dating evidence for early modern humans at Border Cave in South Africa. *Nature,* 344: 537-539.

Hassan F.A. 1978. Archaeological Explorations of the Siwa Oasis Region, Egypt. *Curr. Anthropol.,* 19: 146-148.

Henshilwood C.S., d'Errico F., Yates R., Jacobs Z., Tribolo C., Duller G.A., Mercier N., Sealy J.C., Valladas H., Watts I. & Wintle A.G. 2002. Emergence of modern human behavior: Middle Stone Age engravings from South Africa. *Science,* 295: 1278-1280.

Heyer E., Zietkiewicz E., Rochowski A., Yotova V., Puymirat J. & Labuda D. 2001. Phylogenetic and familial estimates of mitochondrial substitution rates: study of control region mutations in deep-rooting pedigrees. *Am. J. Hum. Genet.,* 69: 1113-1126.

Horai S. 1995. Evolution and the origins of man: clues from complete sequences of hominoid mitochondrial DNA. *Southeast Asian J. Trop. Med. Public Health*, 26: 146-154.

Howell N., Elson J.L., Turnbull D.M. & Herrnstadt C. 2004. African Haplogroup L mtDNA Sequences Show Violations of Clock-Like Evolution. *Mol. Biol. Evol.*, 21: 1843-1854.

Howell N., Smejkal C.B., Mackey D.A., Chinnery P.F., Turnbull D.M. & Herrnstadt C. 2003. The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. *Am. J. Hum. Genet.,* 72: 659-670.

Hünemeier T., Carvalho C., Marrero A.R., Salzano F.M., Junho Pena S.D. & Bortolini M.C. 2007. Niger-Congo speaking populations and the formation of the Brazilian gene pool: mtDNA and Y-chromosome data. *Am. J. Phys. Anthropol.,* 133: 854-867.

Ingman M. & Gyllensten U. 2003. Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. *Genome Res.,* 13: 1600-1606.

Ingman M. & Gyllensten U. 2006. mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences. *Nucleic Acids Res.,* 34: D749-751.

Ingman M., Kaessmann H., Pääbo S. & Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature,* 408: 708-713.

Jackson B.A., Wilson J.L., Kirbah S., Sidney S.S., Rosenberger J., Bassie L., Alie J.A., McLean D.C., Garvey W.T. & Ely B. 2005. Mitochondrial DNA genetic diversity among four ethnic groups in Sierra Leone. *Am. J. Phys. Anthropol.,* 128: 156-163.

Johnson M.J., Wallace D.C., Ferris S.D., Rattazzi M.C. & Cavalli-Sforza L.L. 1983. Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns. *J. Mol. Evol.,* 19: 255-271.

Jorde L.B., Bamshad M.J., Watkins W.S., Zenger R., Fraley A.E., Krakowiak P.A., Carpenter K.D., Soodyall H., Jenkins T. & Rogers A.R. 1995. Origins and affinities of modern humans:

a comparison of mitochondrial and nuclear genetic data. *Am. J. Hum. Genet.,* 57: 523-538.

Kivisild T., Bamshad M.J., Kaldma K., Metspalu M., Metspalu E., Reidla M., Laos S., Parik J., Watkins W.S., Dixon M.E., Papiha S.S., Mastana S.S., Mir M.R., Ferak V. & Villems R. 1999. Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr. Biol.,* 9: 1331-1334.

Kivisild T., Reidla M., Metspalu E., Rosa A., Brehm A., Pennarun E., Parik J., Geberhiwot T., U.S.A.nga E. & Villems R. 2004. Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. *Am. J. Hum. Genet*., 75: 752-770.

Kivisild T., Rootsi S., Metspalu M., Mastana S., Kaldma K., Parik J., Metspalu E., Adojaan M., Tolk H.V., Stepanov V., G"lge M., U.S.A.nga E., Papiha S.S., Cinnioglu C., King R., Cavalli-Sforza L., Underhill P.A. & Villems R. 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am. J. Hum. Genet*., 72: 313-332.

Kivisild T., Shen P., Wall D.P., Do B., Sung R., Davis K., Passarino G., Underhill P.A., Scharfe C., Torroni A., Scozzari R., Modiano D., Coppa A., de K.P., Feldman M., Cavalli-Sforza L.L. & Oefner P.J. 2006. The role of selection in the evolution of human mitochondrial genomes. *Genetics,* 172: 373-387.

Kivisild T., Tolk H.V., Parik J., Wang Y., Papiha S., Bandelt H.J. & Villems R. 2002. The emerging limbs and twigs of the East Asian mtDNA tree. *Mol. Biol. Evol.,* 19: 1737-1751.

Klein H.S., 2002. *As origens africanas dos escravos brasileiros.* Homo brasilis. *Aspectos genéticos, linguisticos, históricos e socioantropológicos da formação do povo brasileiro,* pp. 93-112. FUNPEC Editora, Ribeirão Preto.

Knight A., Underhill P.A., Mortensen H.M., Zhivotovsky L.A., Lin A.A., Henn B.M., Louis D., Ruhlen M. & Mountain J.L. 2003. African Y chromosome and mtDNA divergence provides insight into the history of click languages. *Curr. Biol.*,13: 464-473.

Kong Q.P., Yao Y.G., Sun C., Bandelt H.J., Zhu C.L. & Zhang Y.P. 2003. Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. *Am. J. Hum. Genet.,* 73: 671-676.

Krings M., Salem A.E., Bauer K., Geisert H., Malek A.K., Chaix L., Simon C., Welsby D., Di Rienzo A., Utermann G., Sajantila A., Pääbo S. & Stoneking M. 1999. mtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration? *Am. J. Hum. Genet.,* 64: 1166-1176.

Kujanová M., Pereira L., Fernandes V., Pereira J.B. & Cerny V. 2009. Near eastern neolithic genetic input in a small oasis of the Egyptian Western Desert. *Am. J. Phys. Anthropol.,* 140: 336-346.

Lahr M. & Foley R. 1994. Multiple dispersals and modern human origins. *Evol. Anthropol.,* 3: 48-60.

Lahr M.M. & Foley R.A. 1998. Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution. *Am. J. Phys. Anthropol.,* 41:137-176.

Lewin P.K. 1987. A unique ancient Egyptian mummified head, demonstrating removal of the brain from the foramen magnum. *Paleopathol. Newsl.,* 57, 12-13.

Loogväli E.L., Roostalu U., Malyarchuk B.A., Derenko M.V., Kivisild T., Metspalu E., Tambets K., Reidla M., Tolk H.V., Parik J., Pennarun E., Laos S., Lunkina A., Golubenko M., Barac L., Pericic M., Balanovsky O.P., GU.S.A.r V., Khusnutdinova E.K., Stepanov V., Puzyrev V., Rudan P., Balanovska E.V., Grechanina E., Richard C., Moisan J.P., Chaventre A., Anagnou N.P., Pappa K.I., Michalodimitrakis E.N., Claustres M., Golge M., Mikerezi I., U.S.A.nga E. & Villems R. 2004. Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Mol. Biol. Evol.,* 21: 2012-2021.

Maca-Meyer N., González A.M., Larruga J.M., Flores C. & Cabrera V.M. 2001. Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet.,* 2: 13.

Maca-Meyer N., González A.M., Pestano J., Flores C., Larruga J.M. & Cabrera V.M. 2003. Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography. *BMC Genet.,* 4: 15.

Macaulay V., Hill C., Achilli A., Rengo C., Clarke D., Meehan W., Blackburn J., Semino O., Scozzari

R., Cruciani F., Taha A., Shaari N.K., Raja J.M., Ismail P., Zainuddin Z., Goodwin W., Bulbeck D., Bandelt H.J., Oppenheimer S., Torroni A. & Richards M. 2005. Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science,* 308: 1034-1036.

Macaulay V.A., Richards M.B., Forster P., Bendall K.E., Watson E., Sykes B. & Bandelt H.J. 1997. mtDNA mutation rates--no need to panic. *Am. J. Hum. Genet.,* 61: 983-990.

Macaulay V.A., Richards M.B., Hickey E., Vega E., Cruciani F., Guida V., Scozzari R., Bonn,-Tamir B., Sykes B. & Torroni A. 1999. The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am. J. Hum. Genet.,* 64: 232-249.

McDougall I., Brown F.H. & Fleagle J.G. 2005. Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature,* 433: 733-736.

McMillin J.A., 2004. *The final victims: foreign slave trade to North America,* pp. 1783-1810. University of South Carolina Press, Columbia.

Mellars P. 2006. Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model. *Proc. Natl. Acad. Sci U.S.A.,* 103: 9381-9386.

Mellars P. & Crow T.J., 2002. *The Speciation of Modern* Homo sapiens, pp. 31-47. British Academy, London, U.K..

Mercader J. & Marti R., 2003. *The Middle Stone Age occupation of Atlantic central Africa: new evidence from Equatorial Guinea and Cameroon*, pp. 93-118. Under the Canopy Rutgers University Press, New Brunswick.

Meyer S., Weiss G. & von Haeseler A. 1999. Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics,* 152: 1103-1110.

Mishmar D., Ruiz-Pesini E., Golik P., Macaulay V., Clark A.G., Hosseini S., Brandon M., Easley K., Chen E., Brown M.D., Sukernik R.I., Olckers A. & Wallace D.C. 2003. Natural selection shaped regional mtDNA variation in humans. *Proc. Natl. Acad. Sci. U.S.A.,* 100: 171-176.

Muzzolini A., T.Shaw & et al., 1993. *The emergence of a food-producing economy in the Sahara,* *The archaeology of Africa: food, metals and towns,* pp. 227-239. Routledge, London.

Newman J.L., 1995. *The peopling of Africa*. Yale University Press, New Haven, CT.

Olivieri A., Achilli A., Pala M., Battaglia V., Fornarino S., Al-Zahery N., Scozzari R., Cruciani F., Behar D.M., Dugoujon J.M., Coudray C., Santachiara-Benerecetti A.S., Semino O., Bandelt H.J. & Torroni A. 2006. The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa. *Science,* 314: 1767-1770.

Ottoni C., Martinez-Labarga C., Loogväli E.L., Pennarun E., Achilli A., De Angelis F., Trucchi E., Contini I., Biondi G. & Rickards O. 2009. First genetic insight into Libyan Tuaregs: a maternal perspective. *Ann. Hum. Genet.,* 73: 438-448.

Pereira L., Freitas F., Fernandes V., Pereira J.B., Costa M.D., Costa S., Maximo V., Macaulay V., Rocha R. & Samuels D.C. 2009. The diversity present in 5140 human mitochondrial genomes. *Am. J. Hum. Genet.,* 84: 628-640.

Pereira L., Gusmão L., Alves C., Amorim A. & Prata M.J. 2002. Bantu and European Y-lineages in sub-Saharan Africa. *Ann. Hum. Genet.,* 66: 369-378.

Pereira L., Macaulay V., Torroni A., Scozzari R., Prata M.J. & Amorim A. 2001. Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Ann. Hum.Genet.,* 65: 439-458.

Pereira L., Richards M., Goios A., Alonso A., Albarran C., Garcia O., Behar D.M., Golge M., Hatina J., Al-Gazali L., Bradley D.G., Macaulay V. & Amorim A. 2005. High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. *Genome Res.,* 15: 19-24.

Pereira L., Richards M., Goios A., Alonso A., Albarran C., Garcia O., Behar D.M., Golge M., Hatina J., Al-Gazali L., Bradley D.G., Macaulay V. & Amorim A. 2006. Evaluating the forensic informativeness of mtDNA haplogroup H subtyping on a Eurasian scale. *Forensic Sci. Int.,* 159: 43-50.

Phillipson D., 1993. *African Archaeology*. Cambridge University Press, Cambridge.

Plaza S., Salas A., Calafell F., Corte-Real F., Bertranpetit J., Carracedo A. & Comas D. 2004. Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. *Hum. Genet.*, 115: 439-447.

Quintana-Murci L., Chaix R., Wells S., Behar D., Sayar H., Scozzari R., Rengo C., Al-Zahery N., Semino O., Santachiara-Benerecetti A.S., Coppa A., Ayub Q., Mohyuddin A., Tyler-Smith C., Mehdi Q., Torroni A. & McElreavey K. 2004. Where West meets East: The complex mtDNA landscape of the Southwest and Central Asian corridor. *Am. J. Hum. Genet.*, 74: 827-845.

Quintana-Murci L., Quach H., Harmant C., Luca F., Massonnet B., Patin E., Sica L., Mouguiama-Daouda P., Comas D., Tzur S., Balanovsky O., Kidd K.K., Kidd J.R., van der Veen L., Hombert J.M., Gessain A., Verdu P., Froment A., Bahuchet S., Heyer E., Dausset J., Salas A. & Behar D.M. 2008. Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. *Proc. Natl. Acad. Sci. U.S.A.,* 105: 1596-1601.

Quintana-Murci L., Semino O., Bandelt H.J., Passarino G., McElreavey K. & Santachiara-Benerecetti A.S. 1999. Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nat. Genet.,* 23: 437-441.

Rando J.C., Pinto F., Gonzalez A.M., Hernandez M., Larruga J.M., Cabrera V.M. & Bandelt H.J. 1998. Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. *Ann. Hum. Genet.,* 62: 531-550.

Reidla M., Kivisild T., Metspalu E., Kaldma K., Tambets K., Tolk H.V., Parik J., Loogväli E.L., Derenko M., Malyarchuk B., Bermisheva M., Zhadanov S., Pennarun E., Gubina M., Golubenko M., Damba L., Fedorova S., GU.S.A.r V., Grechanina E., Mikerezi I., Moisan J.P., Chaventre A., Khusnutdinova E., Osipova L., Stepanov V., Voevoda M., Achilli A., Rengo C., Rickards O., De Stefano G.F., Papiha S., Beckman L., Janicijevic B., Rudan P., Anagnou N., Michalodimitrakis E., Koziel S., U.S.A.nga E., Geberhiwot T., Herrnstadt C., Howell N., Torroni A. & Villems R. 2003. Origin and Diffusion of mtDNA Haplogroup X. *Am. J. Hum. Genet.,* 73: 1178-1190.

Richards M., Macaulay V., Hickey E., Vega E., Sykes B., Guida V., Rengo C., Sellitto D., Cruciani F., Kivisild T., Villems R., Thomas M., Rychkov S., Rychkov O., Rychkov Y., Golge M., Dimitrov D., Hill E., Bradley D., Romano V., Cali F., Vona G., Demaine A., Papiha S., Triantaphyllidis C. & Stefanescu G. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.,* 67: 1251-1276.

Richards M., Rengo C., Cruciani F., Gratrix F., Wilson J.F., Scozzari R., Macaulay V. & Torroni A. 2003. Extensive female-mediated gene flow from sub-Saharan Africa into near eastern Arab populations. *Am. J. Hum. Genet.,* 72: 1058-1064.

Richards M.B., Macaulay V.A., Bandelt H.J. & Sykes B.C. 1998. Phylogeography of mitochondrial DNA in western Europe. *Ann. Hum. Genet.,* 62: 241-260.

Rightmire G.P., 1989. *Middle Stone Age humans from eastern and southern Africa, The Human Revolution.* Princeton University Press, Princeton.

Rightmire G.P. & Deacon H.J. 2001. New human teeth from Middle Stone Age deposits at Klasies River, South Africa. *J. Hum. Evol.,* 41: 535-544.

Rightmire G.P., Deacon H.J., Schwartz J.H. & Tattersall I. 2006. Human foot bones from Klasies River main site, South Africa. *J. Hum. Evol.,* 50: 96-103.

Roostalu U., Kutuev I., Loogväli E.L., Metspalu E., Tambets K., Reidla M., Khusnutdinova E.K., U.S.A.nga E., Kivisild T. & Villems R. 2007. Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: the Near Eastern and Caucasian perspective. *Mol. Biol. Evol.,* 24: 436-448.

Rosa A., Brehm A., Kivisild T., Metspalu E. & Villems R. 2004. MtDNA profile of West

Africa Guineans: towards a better understanding of the Senegambia region. *Ann. Hum. Genet.,* 68: 340-352.

Ruiz-Pesini E., Lott M.T., Procaccio V., Poole J.C., Brandon M.C., Mishmar D., Yi C., Kreuziger J., Baldi P. & Wallace D.C. 2007. An enhanced MITOMAP with a global mtDNA mutational phylogeny. *Nucleic Acids Res.,* 35: D823-828.

Salas A., Richards M., De la Fe T., Lareu M.V., Sobrino B., Sanchez-Diz P., Macaulay V. & Carracedo A. 2002. The making of the African mtDNA landscape. *Am. J. Hum. Genet.,* 71: 1082-1111.

Salas A., Richards M., Lareu M.V., Scozzari R., Coppa A., Torroni A., Macaulay V. & Carracedo A. 2004. The African diaspora: mitochondrial DNA and the Atlantic slave trade. *Am. J. Hum. Genet.,* 74: 454-465.

Saunier J.L., Irwin J.A., Strouss K.M., Ragab H., Sturk K.A. & Parsons T.J. 2009. Mitochondrial control region sequences from an Egyptian population sample. *Forensic Sci. Int. Genet*., 3: 97-103.

Scozzari R., Torroni A., Semino O., Sirugo G., Brega A. & Santachiara-Benerecetti A.S. 1988. Genetic studies on the Senegal population. I. Mitochondrial DNA polymorphisms. *Am. J. Hum. Genet.,* 43: 534-544.

Sigurgardottir S., Helgason A., Gulcher J.R., Stefansson K. & Donnelly P. 2000. The mutation rate in the human mtDNA control region. *Am. J. Hum. Genet.,* 66: 1599-1609.

Silva W.A., Bortolini M.C., Schneider M.P., Marrero A., Elion J., Krishnamoorthy R. & Zago M.A. 2006. MtDNA haplogroup analysis of black Brazilian and sub-Saharan populations: implications for the Atlantic slave trade. *Hum. Biol.,* 78: 29-41.

Soares P., Ermini L., Thomson N., Mormina M., Rito T., Röhl A., Salas A., Oppenheimer S., Macaulay V. & Richards M.B. 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am. J. Hum. Genet.,* 84: 740-759.

Soodyall H. & Jenkins T. 1992. Mitochondrial DNA polymorphisms in Khoisan populations from southern Africa. *Ann. Hum. Genet.,* 56: 315-324.

Soodyall H. & Jenkins T. 1993. Mitochondrial DNA polymorphisms in Negroid populations from Namibia: new light on the origins of the Dama, Herero and Ambo. *Ann. Hum. Biol.,* 20: 477-485.

Soodyall H., Vigilant L., Hill A.V., Stoneking M. & Jenkins T. 1996. mtDNA control-region sequence variation suggests multiple independent origins of an "Asian-specific" 9-bp deletion in sub-Saharan Africans. *Am. J. Hum. Genet.,* 58: 595-608.

Stevanovitch A., Gilles A., Bouzaid E., Kefi R., Paris F., Gayraud R.P., Spadoni J.L., El-Chenawi F. & Beraud-Colomb E. 2004. Mitochondrial DNA Sequence Diversity in a Sedentary Population from Egypt. *Ann. Hum. Genet.,* 68: 23-39.

Stringer C. 2000. Coasting out of Africa. *Nature,* 405: 24-25.

Stringer C. 2003. Human evolution: Out of Ethiopia. *Nature,* 423: 692-693.

Stringer C. & Andrews P. 1988. Genetic and fossil evidence for the origin of modern humans. *Science,* 239: 1263-1268.

Sun C., Kong Q.P., Palanichamy M.G., Agrawal S., Bandelt H.J., Yao Y.G., Khan F., Zhu C.L., Chaudhuri T.K. & Zhang Y.P. 2006. The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. *Mol. Biol. Evol.,* 23: 683-690.

Tambets K., Kivisild T., Metspalu E., Parik J., Kaldma K., Laos S., Tolk H.V., Gölge M., Demirtas H., Geberhiwot T., Papiha S.S., de Stefano G.F., Villems R., Renfrew C. & Boyle K., 2000. *The topology of the maternal lineages of the Anatolian and Trans-Caucasus populations and the peopling of the Europe: some preliminary considerations, Archaeogenetics: DNA and the population prehistory of Europe,* pp. 219-235. Cambridge University Press, Cambridge.

Thomas H., 1998. *The slave trade—the history of the Atlantic slave trade,* pp. 1440-1840. Macmillan, London.

Tishkoff S.A., Gonder M.K., Henn B.M., Mortensen H., Knight A., Gignoux C., Fernandopulle N., Lema G., Nyambo T.B., Ramakrishnan U., Reed F.A. & Mountain J.L. 2007. History of

click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Mol. Biol. Evol.,* 24: 2180-2195.

Torroni A., Achilli A., Macaulay V., Richards M. & Bandelt H.J. 2006. Harvesting the fruit of the human mtDNA tree. *Trends Genet.,* 22: 339-345.

Torroni A., Bandelt H.J., Macaulay V., Richards M., Cruciani F., Rengo C., Martinez-Cabrera V., Villems R., Kivisild T., Metspalu E., Parik J., Tolk H.V., Tambets K., Forster P., Karger B., Francalacci P., Rudan P., Janicijevic B., Rickards O., Savontaus M.L., Huoponen K., Laitinen V., Koivumaki S., Sykes B., Hickey E., Novelletto A., Moral P., Sellitto D., Coppa A., Al-Zaheri N., Santachiara-Benerecetti A.S., Semino O. & Scozzari R. 2001a. A signal, from human mtDNA, of postglacial recolonization in Europe. *Am. J. Hum. Genet.,* 69: 844-852.

Torroni A., Rengo C., Guida V., Cruciani F., Sellitto D., Coppa A., Calderon F.L., Simionati B., Valle G., Richards M., Macaulay V. & Scozzari R. 2001b. Do the four clades of the mtDNA haplogroup L2 evolve at different rates? *Am. J. Hum. Genet.,* 69: 1348-1356.

Torroni A., Huoponen K., Francalacci P., Petrozzi M., Morelli L., Scozzari R., Obinu D., Savontaus M.L. & Wallace D.C. 1996. Classification of European mtDNAs from an analysis of three European populations. *Genetics,* 144: 1835-1850.

Torroni A., Schurr T.G., Cabell M.F., Brown M.D., Neel J.V., Larsen M., Smith D.G., Vullo C.M. & Wallace D.C. 1993. Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am. J. Hum. Genet.,* 53: 563-590.

Torroni A., Schurr T.G., Yang C.C., Szathmary E.J., Williams R.C., Schanfield M.S., Troup G.A., Knowler W.C., Lawrence D.N., Weiss K.M. & et al. 1992. Native American mitochondrial DNA analysis indicates that the Amerind and the Nadene populations were founded by two independent migrations. *Genetics,* 130: 153-162.

Trovoada M.J., Pereira L., Gusmão L., Abade A., Amorim A. & Prata M.J. 2004. Pattern of mtDNA variation in three populations from Sao Tome e Principe. *Ann. Hum. Genet.,* 68: 40-54.

Underhill P.A. & Kivisild T. 2007. Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. *Ann. Rev. Genet.,* 41:539-64.

van Oven M. & Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.,* 30: E386-394.

Vigilant L., Stoneking M., Harpending H., Hawkes K. & Wilson A.C. 1991. African populations and the evolution of human mitochondrial DNA. *Science,* 253: 1503-1507.

Watson E., Forster P., Richards M. & Bandelt H.J. 1997. Mitochondrial footprints of human expansions in Africa. *Am. J. Hum. Genet.,* 61: 691-704.

White T.D., Asfaw B., DeGusta D., Gilbert H., Richards G.D., Suwa G. & Howell F.C. 2003. Pleistocene Homo sapiens from Middle Awash, Ethiopia. *Nature,* 423: 742-747.

Wood E.T., Stover D.A., Ehret C., Destro-Bisol G., Spedini G., McLeod H., Louie L., Bamshad M., Strassmann B.I., Soodyall H. & Hammer M.F. 2005. Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. *Eur. J. Hum. Genet.,* 13: 867-876.

Associate Editor, Cristian Capelli

Appendix - Geographic compilation of populations included in the present study and used to generate frequency cline maps. Geographic origin of populations (or sampling localities), ethnic and linguistic affiliation are shown. The linguistic classification follows pertinent information from www.ethnologue.com.

GEOGRAPHIC REGION/ ETHNIC GROUP		POPULATION CODE	LINGUISTIC AFFINITIES	REFERENCE	N
Northwest Africa					
Mauritania	Mauritanians	1	Mixed	Rando *et al*. 1998, González *et al*. 2006	62
West Sahara	Saharawis	2	Afro-Asiatic/Semitic	Rando *et al*. 1998	25
Morocco	Arabs	3	Afro-Asiatic/Semitic	Rando *et al*. 1998	33
	Berbers	4	Afro-Asiatic/Berber	Rando *et al*. 1998	60
Tunisia	Zriba	5	Afro-Asiatic/Semitic	Cherni *et al*. 2009	35
	Kesra	6	Afro-Asiatic/Semitic	Cherni *et al*. 2009	43
	Testour	7	Afro-Asiatic/Semitic	Cherni *et al*. 2009	50
	Slouguia	8	Afro-Asiatic/Semitic	Cherni *et al*. 2009	26
	El Alia	9	Afro-Asiatic/Semitic	Cherni *et al*. 2009	47
	Qalaat El Andalo	10	Afro-Asiatic/Semitic	Cherni *et al*. 2009	29
	Tunis	11	Afro-Asiatic/Semitic	Cherni *et al*. 2009	49
	Skira	12	Afro-Asiatic/Semitic	Cherni *et al*. 2009	20
Libya	Tuaregs	13	Afro-Asiatic/Berber	Ottoni *et al*. 2009	129
East Africa					
Egypt	Egyptians	14	Afro-Asiatic	Saunier *et al*. 2009	263
	Gura	15	Afro-Asiatic	Stevanovich *et al*. 2003	30
Sudan	Nubians	16	Nilo-Saharan/Eastern	Krings *et al*. 1999	80
West Africa					
Cape Verde	Cape Verdeans	17	Creole	Brehm *et al*. 2002	292
Senegal	Senegalese	18	Mixed	Rando *et al*. 1998	50
	Wolof	19	Niger-Congo/ Atlantic-Wolof	Rando *et al*. 1998	48
	Serer	20	Niger-Congo/ Atlantic-Serer	Rando *et al*. 1998	23
Mali	Tuareg	21	Afro-Asiatic/Berber	Watson *et al*. 1996	23

Appendix - Continued

GEOGRAPHIC REGION/ ETHNIC GROUP		POPULATION CODE	LINGUISTIC AFFINITIES	REFERENCE	N
Mali (continued)	Bambara	22	Niger-Congo/ Manding-East	González *et al*. 2006, Ely *et al*. 2006	75
	Malinke	23	Niger-Congo/ Manding-West	González *et al*. 2006, Ely *et al*. 2006	91
Guinea-Bissau	Felupe-Djola	24	Niger-Congo/Atlantic-Bak	Rosa *et al*. 2004	50
	Bijagós	25	Niger-Congo/Atlantic-Bijagó	Rosa *et al*. 2004	22
	Balanta	26	Niger-Congo/Atlantic-Bak	Rosa *et al*. 2004	62
	Papel	27	Niger-Congo/Atlantic-Bak	Rosa *et al*. 2004	77
	Fulbe	28	Niger-Congo /Atlantic-Fulani	Rosa *et al*. 2004	77
	Mandenka	29	Niger-Congo/ Manding-West	Rosa *et al*. 2004	58
	Nalú	30	Niger-Congo/Atlantic-Nalú	Rosa *et al*. 2004	26
Sierra Lione	Mende	31	Niger-Congo/ Manding-Mende-Loko	Jackson *et al*. 2005	58
	Limba	32	Niger-Congo/ Atlantic-Limba	Jackson *et al*. 2005	65
	Temne	33	Niger-Congo/ Atlantic-Temne	Jackson *et al*. 2005	120
	Loko	34	Niger-Congo/ Manding-Mende-Loko	Jackson *et al*. 2005	30
Burkina-Faso	Fulani Banfora	35	Niger-Congo/ Atlantic-Fulani	Cerny *et al*. 2006	50
	Fulani Tindangou	36	Niger-Congo/ Atlantic-Fulani	Cerny *et al*. 2006	47
Central Africa					
Niger/ Nigeria	Yoruba	37	Niger-Congo/Yoruboid	Vigilant *et al*. 1990, Watson *et al*. 1996	33
	Fulbe	38	Niger-Congo/ Atlantic-Fulani	Watson *et al*. 1996	60
	Hausa	39	Afro-Asiatic/Chadic	Watson *et al*. 1996	20
Chad	Mandara	40	Afro-Asiatic/Chadic	Coia *et al*. 2005 (and references therein)	37

Appendix - Continued

GEOGRAPHIC REGION/ ETHNIC GROUP		POPULATION CODE	LINGUISTIC AFFINITIES	REFERENCE	N
Chad (continued)	Uldeme	41	Afro-Asiatic/Chadic	Coia *et al*. 2005 (and references therein)	28
	Podokwo	42	Afro-Asiatic/Chadic	Coia *et al*. 2005 (and references therein)	39
	Fulani Bongor	43	Niger-Congo/ Atlantic-Fulani	Cerny *et al*. 2006	49
North Cameroon	Kotoko	44	Afro-Asiatic/Chadic	Cerny *et al*. 2004	18
	Tupuri	45	Niger-Congo/Adamawa	Coia *et al*. 2005 (and references therein)	24
	Daba	46	Afro-Asiatic/Chadic	Coia *et al*. 2005 (and references therein)	15
	Fali	47	Niger-Congo/Adamawa	Coia *et al*. 2005 (and references therein)	35
	Tali	48	Niger-Congo/Adamawa	Coia *et al*. 2005 (and references therein)	18
	Fulbe	49	Niger-Congo/ Atlantic-Fulani	Coia *et al*. 2005 (and references therein)	26
	Hide	50	Afro-Asiatic/Chadic	Cerny *et al*. 2004	23
	Mafa	51	Afro-Asiatic/Chadic	Cerny *et al*. 2004	32
	Masa	52	Afro-Asiatic/Chadic	Cerny *et al*. 2004	31
South Cameroon	Fulani Tcheboua	53	Niger-Congo/ Atlantic-Fulani	Cerny *et al*. 2006	30
	Bakaka	54	Niger-Congo/Bantu	Coia *et al*. 2005 (and references therein)	49
	Bamilike	55	Niger-Congo/Bantu	Coia *et al*. 2005 (and references therein)	43
	Bassa	56	Niger-Congo/Bantu	Coia *et al*. 2005 (and references therein)	46
	Ewondo	57	Niger-Congo/Bantu	Coia *et al*. 2005 (and references therein)	53
	Agriculturalist*	58	Niger-Congo/Bantu	Quintana-Murci *et al*. 2008	983
	Eastern Pygmies*	59	Niger-Congo/Bantu	Quintana-Murci *et al*. 2008	39
Gabon	Western Pygmies*	60	Niger-Congo/Bantu	Quintana-Murci *et al*. 2008	382

Appendix - Continued

GEOGRAPHIC REGION/ ETHNIC GROUP		POPULATION CODE	LINGUISTIC AFFINITIES	REFERENCE	N
Gabon (continued)	Biaka Pygmies	61	Niger-Congo/Bantu	Vigilant *et al*. 1990, Watson *et al*. 1996	17
	Angolares	62	Creole	Trovoada *et al*. 2004	30
C.A.R.	Forros	63	Creole	Trovoada *et al*. 2004	35
S.Tomé and Principe	Tongas	64	Creole	Trovoada *et al*. 2004	39
	Tigrais	65	Afro-Asiatic/Semitic	Kivisild *et al*. 2004	52
East Africa					
	Oromo/Afar	66	Afro-Asiatic/Cushitic	Kivisild *et al*. 2004	49
Ethiopia	Amhara	67	Afro-Asiatic/Semitic	Kivisild *et al*. 2004	117
	Gurage	68	Afro-Asiatic/Semitic	Kivisild *et al*. 2004	21
	Turkana	69	Nilo-Saharan/Nilotic	Watson *et al*. 1996, 1997	36
	Kikuyu	70	Niger-Congo/Bantu	Watson *et al*. 1996, 1997	25
Kenya	Nairobi	71	Mixed	Brandstätter *et al*. 2004	84
	Somalians	72	Afro-Asiatic/Mixed	Watson *et al*. 1996, 1997	27
	Shona	73	Niger-Congo/Bantu	Castri *et al*. 2009	59
Somalia	Hutu	74	Niger-Congo/Bantu	Castri *et al*. 2009	42
Rwanda					
	Mozambicans	75	Mixed	Pereira *et al*. 2001	109
South Africa	Bantu	76	Niger-Congo/Bantu	Salas *et al*. 2002	304
Mozambique	!Kung	77	Khoisan/Northern	Chen *et al*. 2000	43
	Khwe	78	Khoisan/Central	Chen *et al*. 2000	31
South Africa	Mbundu	79	Niger-Congo/Bantu	Plaza *et al*. 2004	44
	Cabinda	80	Niger-Congo/Bantu	Beleza *et al*. 2005	107
Angola	Kuvale	81	Niger-Congo/Bantu	Coelho *et al*. 2009	71
	Ganguela	82	Niger-Congo/Bantu	Coelho *et al*. 2009	27
	Nyaneka-Nkhumbi	83	Niger-Congo/Bantu	Coelho *et al*. 2009	175

Appendix - Continued

GEOGRAPHIC REGION/ ETHNIC GROUP		POPULATION CODE	LINGUISTIC AFFINITIES	REFERENCE	N
Angola (continued)	Ovimbundu	84	Niger-Congo/Bantu	Coelho *et al*. 2009	101
	Others (mixed)	85	Mixed	Coelho *et al*. 2009	45
	Hadza	86	Khoisan	Tishkoff *et al*. 2007	79
	Sandawe	87	Khoisan	Tishkoff *et al*. 2007	82
Tanzania	Sukuma	88	Niger-Congo/Bantu	Tishkoff *et al*. 2007	11
	Turu	89	Nilo-Saharan/Nilotic	Tishkoff *et al*. 2007	30
	Datooga	90	Nilo-Saharan/Nilotic	Tishkoff *et al*. 2007	39
	Burunge	91	Afro-Asiatic/Cushitic	Tishkoff *et al*. 2007	38

* samples with ambiguous assignment for the phylogenetic clades considered in the present review, were not included